

Access DB# 61843

SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: \_\_\_\_\_ Examiner #: \_\_\_\_\_ Date: \_\_\_\_\_  
Art Unit: \_\_\_\_\_ Phone Number 30 \_\_\_\_\_ Serial Number: \_\_\_\_\_  
Mail Box and Bldg/Room Location: \_\_\_\_\_ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

\*\*\*\*\*

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc. if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: \_\_\_\_\_

Inventors (please provide full names): \_\_\_\_\_

Earliest Priority Filing Date: \_\_\_\_\_

*\*For Sequence Searches Only\* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

Jan Delaval  
Reference Librarian  
Biotechnology & Chemical Library  
CM1 1E07 - 703-308-4498  
jan.delaval@uspto.gov

STAFF USE ONLY

	Type of Search	Vendors and cost where applicable
Searcher _____	NA Sequence (#) <input checked="" type="checkbox"/>	STN _____
Searcher Phone #: 44451	AA Sequence (#) <input checked="" type="checkbox"/>	Dialog _____
Searcher Location: _____	Structure (#) _____	Questel/Orbit _____
Date Searcher Picked Up: 4/20/02	Bibliographic _____	Dr. Link _____
Date Completed: 4/22/02	Litigation _____	Lexis/Nexis _____
Searcher Prep & Review Time: _____	Fulltext _____	Sequence Systems <input checked="" type="checkbox"/>
Clerical Prep Time: 10	Patent Family _____	WWW/Internet _____
Online Time: 10	Other _____	Other (specify) _____

BEST AVAILABLE COPY





JOURNAL Patent: WO 0173028-A 1 04-OCT-2001;

GENETICS INSTITUTE, INC. (US)

FEATURES Location/Qualifiers

1..951

/organism="Homo sapiens"

/db\_xref="taxon:9606"

CDs

1..951

/note="unnamed protein product"

/codon\_start=1

/protein\_id="CAD10213.1"

/db\_xref="GI:16508827"

/translation="MASPBHDSPOCMSPITQCTARTQEAQATPAPDLPHPGPDGILD  
THSGLSNSMTRELQYQWONOKRMKHWKLEIPIASARIEEKVKSPVYQIIVD  
TGSFNRHVALEFSEDFALOKALKTFPEEIDVEFRKIDGKFAEEMICERRA  
LOEYLGLYAIRCYRBSREPLDLPLRLPRLRACGLRAGQYPRALLELVLPQEKU  
TAHCPRAAVPACAVLCHRDLPDPAEPFAAGEALQRLCARBSHRYAPLDLMAVPL  
AYALGDFPTLQEPLEESQLPPTPPGTLTKELTVPEYLH"BASE COUNT 183 a 320 c 297 g 151 t  
ORIGINQuery Match 100.0%; Score 951; DB 6; Length 951;  
Best Local Similarity 100.0%; Pred. No. 4 Re-120;

Matches 951; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCAAGTCCAGAGTACCTGCTGTAAGCTGCTGATGAGAGACCATACAGAGTGCACG 60  
 Db 1 ATGGCAAGTCCAGAGTACCTGCTGTAAGCTGCTGATGAGAGACCATACAGAGTGCACG 60  
 QY 61 GCAGGACCTGAGAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 120  
 Db 61 GCAGGACCTGAGAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 120  
 QY 121 GGGGACTTACACACACACAGTACCTGAGTCCCACTCCAGATACACACGCGGAGCTT 180  
 Db 121 GGGGACTTACACACACAGTACCTGAGTCCCACTCCAGATACACACGCGGAGCTT 180  
 QY 181 CAGTAACTATTAATAAATATTAATAAATATTAATAAATATTAATAAATATTAATAAAT 240  
 Db 181 CAGTAACTATTAATAAATATTAATAAATATTAATAAATATTAATAAATATTAATAAAT 240  
 QY 241 GCTTCAGCTGCTGATGAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 300  
 Db 241 GCTTCAGCTGCTGATGAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 300  
 QY 361 ATCCAGACTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 420  
 Db 361 ATCCAGACTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 420  
 QY 421 CTTAGAGAGAGTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 480  
 Db 421 CTTAGAGAGAGTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 480  
 QY 481 CTGAGAGTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 540  
 Db 481 CTGAGAGTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 540  
 QY 541 CTGAGAGTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 600  
 Db 541 CTGAGAGTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 600  
 QY 601 TACCCGCGCGCGCTGAGT 660  
 Db 601 TACCCGCGCGCGCTGAGT 660  
 QY 661 CATTCGCTGCGT 720  
 Db 661 CATTCGCTGCGT 720  
 QY 721 GATTCGCTGCGT 780

Db 721 GATTCGCTGCGT 780  
 QY 781 GATTCGCTGCGT 840  
 Db 781 GATTCGCTGCGT 840  
 QY 841 GATTCGCTGCGT 900  
 Db 841 GATTCGCTGCGT 900  
 QY 901 CTTAGAGAGAGTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 951  
 Db 901 CTTAGAGAGAGTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 951

## RESULT 2

AX259777

LOCUS AX259777 1315 bp DNA linear PAT 26-OCT-2001

DEFINITION Sequence 4 from Patent W00172822.

ACCESSION AX259777

VERSION AX259777.1 GI:16508821

KEYWORDS

SOURCE

ORGANISM

human.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Placentalia; Catarrhini; Homnidae; Homo.

REFERENCE

1

Hugot, P., Thomas, G., Douali, M., Lesage, S. and Chamallard, M.

Genes involved in intestinal inflammatory diseases and use thereof

Patent: WO 0172822 A 4 04-OCT-2001;

Fondation Jean Dausset-Ceph (FR)

Location/Qualifiers

1..1315

/organism="Homo sapiens"

&lt;db\_xref="taxon:9606"

&lt;117..1121

/note="unnamed protein product"

/codon\_start=1

/protein\_id="CAD10213.1"

/translation="HSSRTPGHQDPWFLSEWMSPEHSDPCMSPIITQCTARTQOE  
APATGPDLPHPGPDGILDTHSGLSNSMTRELQYQWONOKRMKHWKLEIPIASAR  
IEERKVKSPVYQIIVDQSGFDNKNKAVLEFRSPDALKLTKTFPEEIDVEFR  
KIDGKFAEEMICERRA LOEYLGLYAIRCYRBSREPLDLPLRLPRLRACGLRAGQYPRALLELVLPQEKU  
TAHCPRAAVPACAVLCHRDLPDPAEPFAAGEALQRLCARBSHRYAPLDLMAVPL  
AYALGDFPTLQEPLEESQLPPTPPGTLTKELTVPEYLH  
ARBSHRYAPLDLMAVPLAYALGDFPTLQEPLEESQLPPTPPGTLTKELTVPEYLH"

BASE COUNT 257 a 413 c 394 g 245 t

ORIGIN

Query Match 99.8%; Score 949.4; DB 6; Length 1315;

Best Local Similarity 99.9%; Pred. No. 1e-119;

Matches 950; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGGCAAGTCCAGAGTACCTGCTGTAAGCTGCTGATGAGAGACCATACAGAGTGCACG 49  
 Db 171 ATGGCAAGTCCAGAGTACCTGCTGTAAGCTGCTGATGAGAGACCATACAGAGTGCACG 230  
 QY 231 GCAAGGACCTGAGAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 290  
 Db 231 GCAAGGACCTGAGAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 290  
 QY 291 GGTGATTTAGACACACAGT 350  
 Db 291 GGTGATTTAGACACACAGT 350  
 QY 351 GGTGATTTAGACACACAGT 410  
 Db 351 GGTGATTTAGACACACAGT 410  
 QY 410 GGTGATTTAGACACACAGT 470  
 Db 410 GGTGATTTAGACACACAGT 470







Dh	72640	GCAGAGAAAGCTTCACCGGCACTGCGCTCGCGCGCGCGCTGCCGCGCTGCT	72581
Gy	702	GCTGGTCACCGCGGACTCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT	761
Dh	72560	GATTGGGACACCGGAGACCTTGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCT	72521
Gy	742	GCAGAGCGCTTGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT	821
Dh	72520	GCAGAGCT	72461
Gy	822	CGCT	881
Dh	72460	CGCT	72401
Gy	882	CGAGCTTCGCT	941
Dh	72400	CGAGCTTCGCT	72341
Gy	942	CCTGCACCTGA 951	
Dh	72340	CCTGCACCTGA 72331	

RESULT	5	1,69773 bp	DNA	1 linear	PRI 11:CA:2023
LOCUS	AC007608				
DEFINITION	Homo sapiens chromosome 16 clone RP11-401P9, complete sequence.				
ACCESSION	AC007608				
VERSION	AC007608.6	GI:18129388			
KEYWORDS	HTG;				
SOURCE	Homo sapiens.				
ORGANISM	Homo sapiens				

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Hammer, D., Lander, E., Zakariasen, S., Smith, H., Horn, L., Cawthon, C., et al. (1997). The Joint Genome Institute. Sequencing of Human Chromosome 16 (bases 1 to 169773).  
Unpublished  
2 (bases 1 to 169773)  
Bruce, D., Mundt, M., Deggett, N., Munk, C., Sanders, E., Robinson, D.

Jones, M., Buckingham, J., Chasteen, L., Thompson, S., Goodin, L.,  
 Bryant, J., Tesmer, J., Meincke, L., Longmire, J., White, S., Tetum, O.,  
 Campbell, C., Fawcett, J., Matthe, M., Bussod, M., Sutherland, R.,  
 McMurtry, K., Han, C. and Deaven, L.  
 Direct Submission  
 Submitted (20-MAY-1999) Center for Human Genome Studies, DOE Joint  
 Genome Program

REFERENCE  
AUTHORS  
3 (bases 1 to 169773)  
DOE Joint Genome Institute.

Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
4 (bases 1 to 169773)

COMMENT On Jan 11, 2002 this sequence version replaced gi:18057072.

**Sequence Quality Assessment:** This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the GenBank flat file format, but are available as part of this entry's ASN.1 file.

-----  
Sequence Quality Assessment:  
This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero.

Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

FEATURES	Location/Qualifiers
source	1. .169773

BASE COUNT	43799 a	45185 c	41101 g	39688 t
ORIGIN				

```

Query Match      70.5%; Score 670; DB 9; Length 169773;
Best Local Similarity 100.0%; Pctd No 2 7e-A2;
Matches 670; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 282 GGGTACCAAAATCATCTCATCCAGATTGGTAGTTTAAACAAACAAGGCGCTCTTGGAA 3411  
Db 37640 GGTTACCAAAATCATCTCATCCAGATTGGTAGTTTAAACAAACAAGGCGCTCTTGGAA 3766

Oy 342 ACGGCGCTATCCGATTCCGGAGCGTCCGASAAAGCGTGTGTAAGACGCTTCAGGAGGGA 401  
 Dp 37780 ACGGCGCTATCCGATTCCGGAGCGTCCGASAAAGCGTGTGTAAGACGCTTCAGGAGGGA 37780

QY 402 GATCGAGAGCTGGAGTTTCCCGAGAGCACTGACTGGAACTTCGCTGAGGAGATGAT 461

DJ 37760 GATCGAGAGCTGGAGTTTCCCGAGAGCACTGACTGGAACTTCGCTGAGGAGATGAT 378

QY 462 CTGTGAGCGCTGGGCGGCTGCTTGTGTAGAGATACCTGCTGCTTGTGTACGGCTATCCGCTGCGT 521

Db 37820 CTGTGAGCGCTGGGCGGCTGCTTGTGTAGAGATACCTGCTGCTTGTGTACGGCTATCCGCTGCGT 378

QY	522 GCGCGCTCCGGAGACTTCCTGACATTTCTAAGAGCTTTCG	581
Db	37880 GCGCGCTCCGGAGACTTCCTGACATTTCTAAGAGCTTTCG	379

QY 582 CTGCTGGCGGCCGCAGTACCCGCGCGCCCTTGCAATGTATTTGGCGCGTAAGTAACTGCT 641

Db 37940 CTGCTGCGGCGCCGAGTACCCTGGCGCTGTGAAGTATATTTGGCGCGTAAGTAACTGCT 439

QY 642 GCAGGAGAACCTCACCGGCCCACTGCGCTTGGCGAGATCGCGCTCCGAGGCTGTATGAGCGCTGCT 701

Db 3800 GCAGGAGAACCTCACCGGCCCACTGCGCTTGGCGAGATCGCGCTCCGAGGCTGTATGAGCGCTGCT 340

Qy	702	GCCTGCGCACCGCGACCTCGACACGCGCGCCCGCAGAGGCTTTCGCGCGGAGAGAGGCGCTT	761
Dc	38060	GCCTGCGCACCGCGACCTCGACACGCGCGCGCCCGCAGAGGCTTTCGCGCGGAGAGAGGCGCGCTT	38119

Oy	762	GCAGCGCCTGCAAGATCGGAGAGGCCATGCTACTATATGAGCCCTTGCTGAGAGCCATGCT	832
D6	38120	GCAGCGCCTGCAAGATCGGAGAGGCCATGCTACTATGAGCCCTTGCTGAGAGCCATGCT	38121

OY	822	CCGGCTGGCCTACCGGCTGGGTAAAGCACTTCTGTACTTTTCAGAGCAAGCTTGGAAGGAG	881
Dc	38186	CCGGCTGGCCTACCGGCTGGGTAAAGCACTTCTGTACTTTTCAGAGCAAGCTTGGAAGGAG	442

QY	882	CCAGCTCCGGAGGAGGCTACGCTCGGAGGAGGCTCACTGTGCGAGAGATA	9411
Db	38240	CCAGCTCGGAGGAGGCTCGGAGGAGGCTCACTGTGCGAGAGATA	9533

Qy 942 CCTGCACCTGA 951  
|||||  
Db 38300 CCTGCACCTGA 38309

RESULT 6  
AE395844

LOCUS	AF395844	942 bp	mRNA	linear	ROD 16-JUL-2001
DEFINITION	Mus musculus SNX20 (SNX20) mRNA, complete cds.				
ACCESSION	AF395844				















QY 707 GGCACCGGACCTTCGACCGGCGGAGGCTTCGGGAGGAGGAGGCGCTTCGAC 766  
 Db 943 GATACATAGAGACATAGAGACATAGAGACATAGAGACATAGAGACATAGAGAC 1002  
 QY 767 GCGTCGACGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 826  
 Db 1003 TCGTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1062  
 QY 827 TGGCTACGCGCTGGGAGGAGGAG 849  
 Db 1063 TCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1085

RESULT 12  
 G67944/c  
 LOCUS 667944 270 bp DNA linear STS 31-JUL-2001  
 DEFINITION 1BD1pox-exonc genomic STS on chromosome 16q12.13 Homo sapiens STS  
 accession 667944  
 version 667944.1 GI:13442985  
 KEYWORDS STS  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini, Homnidae, Homo.

REFERENCE  
 AUTHORS Hugot, J.P., Chamaillard, M., Zouali, H., Lesage, S., Cezaire, J.P.,  
 Belaiche, J., Almer, S., Tysk, C., O'Morain, C.A., Gassull, M.,  
 Binder, V., Finkler, Y., Corot, A., Modigliani, P., Laurent-Puig, P.,  
 Gower-Rousseau, C., Macry, J., Colombet, J.F., Sahbatou, M. and  
 Thomas, G.  
 Association of NOD2 leucine-rich repeat variants with  
 susceptibility to Crohn's disease  
 Nature 411 (6837), 599-603 (2001)  
 JOURNAL 21279172  
 MEDLINE 11385576  
 PUBMED  
 COMMENT

Contact: Habib Zouali  
 Inflammatory Bowel Diseases  
 Centre d'Etude du Polymorphisme Humain (CEPH)  
 27, rue Juliette Doda 75010 Paris, France  
 Tel: 00 33 1 53 72 50 20  
 Fax: 00 33 1 53 72 50 58  
 Email: habib@ceph.fr  
 Primer A: GGACGACAGTGAAGAGG  
 Primer B: GCGATCTCCCAAGCC  
 STS size: 270  
 PCR profile:  
 Initial denaturation step of genomic DNA at 95 degrees C for 12 min  
 to  
 activate AmpliTaq Gold,  
 30 cycles: 94 degrees C for 30 sec, annealing for 30 sec at 55  
 degrees C, extension at 72 degrees C for 30 sec; last extension at  
 72 degrees C for 7 min.  
 PCR was carried out on Thermal Cycler (MJ Research, PTC-200) .  
 Protocol:  
 Template (genomic DNA) 100 ng  
 Primer: each 1 uM  
 dNTPs: each 200 uM  
 MgCl2 1.5 mM  
 Taq Polymerase 0.05 units/ul  
 Total Volume 20 ul  
 Polymerase : Taq Gold (Roche Perkin Elmer)  
 Buffer:  
 Tris-HCl 10 mM  
 KCl 50 mM  
 pH 8.3 at 20°C

FEATURES  
 source  
 Unigene EST cluster hs.87280; This STS contains a coding sequence  
 which was identified by screening of Human cDNA library.  
 Location/Qualifiers  
 1..270

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="Genomic STS on chromosome 16q12.13"  
 /note="Genomic STS were identified on Human CEPH-BAC clone  
 (hB87b1c) located on chromosome 16 using shot-gun and  
 sequencing procedures."  
 1..270  
 primer\_bind 1..19  
 primer\_bind complement(254..270)  
 BASE COUNT 59 a 71 c 74 g 66 t  
 ORIGIN

Query Match 16.2%; Score 154.4; DP 11; Length 270;  
 Best Local Similarity 99.4%; P-Val: 7.5e-12;  
 Matches 155; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 129 AGACACACACAGTGGCTGAGCTTCAGCTTCAGATACAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 198  
 Db 200 AGACACACACAGTGGCTGAGCTTCAGCTTCAGATACAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 141  
 QY 189 CTGGCAGACCGAGAAATGCGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 248  
 Db 140 CTGGCAGACCGAGAAATGCGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 81  
 QY 249 TCGCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 284  
 Db 80 TCGCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 45

RESULT 13  
 AK095851  
 LOCUS 3044 bp mRNA linear p15-JUL-2002  
 DEFINITION Homo sapiens cDNA FL179832 f1s, clone HCHON2001099.  
 accession AK095851  
 version AK095851.1 GI:21755194  
 keywords oligo capping; f1s (full insert sequence).  
 SOURCE Homo sapiens chondrocytes (HC) cDNA to mRNA, clone lib:HCHON2  
 clone:HCHON2001099.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini, Homnidae; Homo.

REFERENCE  
 AUTHORS Tanigami, A., Fujiwara, T., Shibahara, T., Goto, Y., Hiraio, M.,  
 Shimizu, F., Wakebe, H., Ono, T., Hishigaki, H., Watanabe, T., Ozaki, K.,  
 Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S.,  
 Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T.,  
 Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M.,  
 Kikuchi, H., Kanda, K., Magatsuma, M., Murakawa, K., Kanehori, K.,  
 Takahashi-Fujii, A., Gehlma, A., Sugiyama, A., Kawakami, B., Suzuki, Y.,  
 Sugano, S., Nagahara, K., Masuko, Y., Nagai, K. and Isogai, T.  
 NEDO human cDNA sequencing project

TITLE Unpublished  
 JOURNAL 2 (bases 1 to 3044)  
 REFERENCE Isogai, T. and Yamamoto, J.  
 AUTHORS Direct Submission  
 JOURNAL Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7  
 Kazusa-Kamatori, Kisarazu, Chiba 252-0812, Japan  
 (E-mail: genomics@n1.co.jp, Tel: 81-438-52 375, Fax: 81 438 52-3986)  
 NEDO human cDNA sequencing project supported by Ministry of  
 Economy, Trade and Industry of Japan; cDNA full insert sequencing:  
 Research Association for Biotechnology (RAB); cDNA library  
 construction: Helix Research Institute (HRI) (supported by Japan  
 Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,  
 HRI, and Biotechnology Center, National Institute of Technology, RAB,  
 Evaluation: clone selection for full insert sequencing: HRI and  
 RAB; annotation: HRI and RAB.  
 Location/Qualifiers  
 1..3044

FEATURES  
 source  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="HCHON2001099"  
 /note="Genomic STS were identified on Human CEPH-BAC clone  
 (hB87b1c) located on chromosome 16 using shot-gun and  
 sequencing procedures."  
 1..270  
 primer\_bind 1..19  
 primer\_bind complement(254..270)  
 BASE COUNT 59 a 71 c 74 g 66 t  
 ORIGIN







GenCore version 5.1.4.g5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: April 20, 2003, 15:50:09 ; Search time 3711 Seconds  
(without alignments)

4150.339 Million cell updates/sec

Title: US-09-816-697A-1

Sequence: 1 atggcaagatccagagatccc tgcgagaataacctgcaactga gct

Scoring table: IDENTITY\_NUC

Gapop 10.0, Gapext 1.0

Searched: 1c154066 seqs, 900774176 residues

Total number of hits satisfying chosen parameters: 32300132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estha:\*  
2: em\_esthum:\*  
3: em\_estlin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vit:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rnd:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	693.8	73.0	934	14	BQ706954 AGENCOURT
2	667.4	70.2	772	14	BQ230734 AGENCOURT
3	623.8	65.6	1437	11	AK018632 Mus muscu
4	565.4	59.5	567	10	AM134842 UT-H-B1-
5	540	56.8	540	9	A176326 w997805.X
6	527.4	55.5	529	10	AM082334 x065f03.X

C 7	491.8	51.7	506	10	AM467091	AM467091 ha09f03.x
C 8	470	49.4	478	9	A1004982	A1004982 ou07a09.x
C 9	461.4	48.5	671	13	B1099756	B1099756 6028494.x
C 10	460	48.4	460	9	A1243205	A1243205 qh40h05.x
C 11	460	48.4	460	10	AM592491	AM592491 hf43c11.x
C 12	460	48.4	461	9	A1524170	A1524170 th10a08.x
C 13	459	48.3	459	9	A1131560	A1131560 qb82d04.x
C 14	455.8	47.9	509	10	AM469821	AM469821 hd45f08.x
C 15	454	47.7	500	9	A1554174	A1554174 tes0d05.x
C 16	437	46.0	437	9	A1017999	A1017999 cu36c01.x
C 17	436.4	45.9	742	13	B1409523	B1409523 602963922
C 18	436	45.8	460	9	AA807226	AA807226 c236g04.s
C 19	435.4	45.8	542	9	AA724066	AA724066 ah86f02.s
C 20	432	45.4	504	13	EA149286	EA149286 TCAPFC11
C 21	431.4	45.4	514	9	AA909178	AA909178 o112a07.s
C 22	428.4	45.0	543	9	AA521424	AA521424 jad3a11.s
C 23	422.4	44.4	426	9	A1582328	A1582328 t566g07.x
C 24	413.4	43.5	542	9	AA207906	AA207906 z577c05.s
C 25	407	42.8	541	13	EM135126	EM135126 TCAPFC10
C 26	401.4	42.2	471	9	AA993970	AA993970 cu41b04.s
C 27	396.6	41.7	409	10	AA971924	AA971924 EST384013
C 28	394.8	41.5	476	9	AA884511	AA884511 al56h02.s
C 29	386	40.6	386	9	A1480235	A1480235 tmc2a01.x
C 30	385	40.5	518	13	EM194142	EM194142 TCAPFC10
C 31	379.2	39.9	529	13	EM149874	EM149874 TCAPFC10
C 32	378.4	39.8	484	9	A1538472	A1538472 td06h05.x
C 33	378	39.7	445	9	AA972085	AA972085 cp76e12.s
C 34	377	39.6	557	9	AA282390	AA282390 z589a11.x
C 35	372.4	39.2	396	9	AA831579	AA831579 oc84f06.s
C 36	369.8	38.9	392	9	AA872215	AA872215 ob78h10.s
C 37	363.2	38.2	626	10	BB645468	BB645468 BB645468
C 38	362.4	38.1	483	13	EM193158	EM193158 TCAPFC10
C 39	348.8	36.7	496	10	AM825558	AM825558 us17h05.y
C 40	339.2	35.7	484	13	AM146883	AM146883 TCAPFC17
C 41	337.6	35.5	632	10	AM701659	AM701659 u488h12.y
C 42	332.4	35.0	435	9	AA978303	AA978303 op46f12.s
C 43	330.6	34.8	670	10	BB637909	BB637909 BB637909
C 44	327.2	34.4	388	9	AA933915	AA933915 on91c11.s
C 45	325.8	34.3	424	9	AA885032	AA885032 amd1e07.s

## ALIGNMENTS

RESULT 1  
LOCUS BQ706954 934 bp mRNA linear EST 16-JUL-2002  
DEFINITION AGENCOURT 8353554 NIH\_MGC\_113 Homo sapiens CDNA clone IMAGE:6279494  
ACCESSION BQ706954  
VERSION BQ706954.1 GI:21845853  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 934)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Dr. Mark Watson  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
File: LNC2469.tcw. C column. 15  
High quality sequence stop: 567.  
Location/Qualifiers 1..934

FEATURES  
source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6279494"
/lab_host="NIH_MGC_113"
/notes="Organ: spleen; Vector: pOT8; Site 1: XhoI; Site 2: EcoRI; CDNA made by oligo-dt priming, directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH MGC Library."
BASE COUNT      191 a      279 g      142 t
ORIGIN
Query Match      73.0%; Score 693.8; DB 14; Length 934;
Best Local Similarity 90.6%; Pred. No. 1.1e-130;
Matches 77; Conservative 0; Mismatches 77; Indels 3; Gaps 3,
1  ATGCGAAGTCCAGACGACCTGGAGAGCCCTGGTCATGAGGACCATTAACCGAGTGCAG 60
Db  82  ATGCGAAGTCCAGACGACCTGGAGAGCCCTGGTCATGAGGACCATTAACCGAGTGCAG 141
QY  61  GCAAGGACCAAGAGGAAAGATGAGCAATGAGCCGACCTCCGACCCAGACCTGAC 120
Db  142  GCAAGGACCAAGAGGAAAGATGAGCAATGAGCCGACCTCCGACCCAGACCTGAC 201
QY  121  GGGCACTTAAATACACATCTGAGAGTCCGACCTCCGACCTCCGACCTCCGACCT 180
Db  202  GAGGATTTAAATACACATCTGAGAGTCCGACCTCCGACCTCCGACCTCCGACCT 261
QY  181  CAGCAATTAATGAGAAATGAGAAATGAGAAATGAGAAATGAGAAATGAGAAATG 240
Db  262  CAGCAATTAATGAGAAATGAGAAATGAGAAATGAGAAATGAGAAATGAGAAATG 321
QY  241  GCTTCAGCTGATGATGAGAGGAGAAAGTCTTAAGTTTGTGTTGCAATATCTGTC 300
Db  322  GCTTCAGCTGATGATGAGAGGAGAAAGTCTTAAGTTTGTGTTGCAATATCTGTC 381
QY  301  ATTCAATTAATGAGAAATGAGAAATGAGAAATGAGAAATGAGAAATGAGAAATG 360
Db  382  ATTCAATTAATGAGAAATGAGAAATGAGAAATGAGAAATGAGAAATGAGAAATG 441
QY  361  GGTAAATTAATGAGAAATGAGAAATGAGAAATGAGAAATGAGAAATGAGAAATG 420
Db  442  GGTAAATTAATGAGAAATGAGAAATGAGAAATGAGAAATGAGAAATGAGAAATG 501
QY  421  CCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 480
Db  502  CCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 561
QY  481  CCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 539
Db  562  CCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 621
QY  540  CCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 599
Db  622  CCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 681
QY  600  GTACCCGCGGCGCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 659
Db  682  GTACCCGCGGCGCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 741
QY  660  GTACCCGCGGCGCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 718
Db  742  GTACCCGCGGCGCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 801
QY  719  TTTAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 777
Db  802  TTTAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 861
QY  778  CTTAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 837

```

```

Db  862  GAGGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 921
QY  838  CTGGGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 850
Db  922  CGGCTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 934

RESULT 2
BQ230734
LOCUS
DEFINITION
AGENCOURT 7565875 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6279494
5' mRNA sequence.
ACCESSION
BQ230734
VERSION
BQ230734.1 GI:20412134
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 772)
NIH-MGC http://mgi.nhl.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: rgs@nsl.nhl.nih.gov
Tissue procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Aided by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLM1279 Row: n Column: 09
High quality sequence stop: 643.
Location/Qualifiers
1. 772
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6041672"
/lab_host="NIH_MGC_92"
/tissue_type="embryonal carcinoma, cell line"
/notes="Organ: testis; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; cloned unidirectionally; oligo-dt primed.
Average insert size 2.5 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT      164 a      249 c      231 g      128 t
ORIGIN
Query Match      70.2%; Score 667.4; DB 14; Length 772;
Best Local Similarity 99.9%; Pred. No. 2.4e-125;
Matches 668; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
1  ATGCGAAGTCCAGACGACCTGGAGAGCCCTGGTCATGAGGACCATTAACCGAGTGCAG 60
Db  104  ATGCGAAGTCCAGACGACCTGGAGAGCCCTGGTCATGAGGACCATTAACCGAGTGCAG 163
QY  61  GCAAGGACCAAGAGGAAAGATGAGCAATGAGCCGACCTCCGACCCAGACCTGAC 120
Db  164  GCAAGGACCAAGAGGAAAGATGAGCAATGAGCCGACCTCCGACCCAGACCTGAC 223
QY  121  GGGCACTTAAATACACATCTGAGAGTCCGACCTCCGACCTCCGACCTCCGACCT 180
Db  224  GGGCACTTAAATACACATCTGAGAGTCCGACCTCCGACCTCCGACCTCCGACCT 283
QY  181  CAGCAATTAATGAGAAATGAGAAATGAGAAATGAGAAATGAGAAATGAGAAATG 240
Db  284  CAGCAATTAATGAGAAATGAGAAATGAGAAATGAGAAATGAGAAATGAGAAATG 343
QY  241  GCTTCAGCTGATGATGAGAGGAGAAAGTCTTAAGTTTGTGTTGCAATATCTGTC 300

```

Db 344 GCTTCAGCTCCGATCGAGAGAGAAAGCTCTTAAGTTTGTGTACCAATCATCTGTC 403

QY 301 ATCCAGACTGGAGCTTTGACACAAAGCCGCTCTTGAAGCCGCTATTCGAGCTTC 360

Db 404 ATCCAGACTGGAGCTTTGACACAAAGCCGCTCTTGAAGCCGCTATTCGAGCTTC 463

QY 361 GCGAAGCTTCGAGAAAGGCTCTGTAAGACGCTGAGGAGGAGATCGAAGCCGCTGATTT 420

Db 464 GCGAAGCTTCGAGAAAGGCTCTGTAAGACGCTGAGGAGGAGATCGAAGCCGCTGATTT 523

QY 421 GCGAAGCTTCGAGAAAGGCTCTGTAAGACGCTGAGGAGGAGATCGAAGCCGCTGATTT 480

Db 524 GCGAAGCTTCGAGAAAGGCTCTGTAAGACGCTGAGGAGGAGATCGAAGCCGCTGATTT 583

QY 481 CTGACAGAGATACCTGAGGCTCTGATGAGGAGATGATCTGTGAGCGCTGAGCGCC 540

Db 584 CTGACAGAGATACCTGAGGCTCTGATGAGGAGATGATCTGTGAGCGCGCTGAGCGCC 643

QY 541 CTGACAGATCTGAGGAGGCTGAGGAGGAGGCTGAGGAGGCTGAGGAGGAGGAGGAG 600

Db 644 CTGACATCTCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 703

QY 601 TACCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 660

Db 704 TACCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 763

QY 661 CACTGCGCT 669

Db 764 CACTGCGCT 772

RESULT 3  
AK018632 1437 bp mRNA linear HTC 19-JAN-2002

LOCUS  
DEFINITION  
Mus musculus adult male cecum cDNA, RIKEN full-length enriched library, clone:9130017C17;PX domain containing protein, full insert sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

1 Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)

2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)

3 Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Katsunari, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujisake, S., Inoue, K., Togawa, Y., Izawa, M., Ohata, E., Watanabe, M., Okazaki, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunura, S., Kawai, J., RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer  
JOURNAL  
MEDLINE  
PUBMED  
11076861

# REFERENCE AUTHORS

4 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Aizawa, K., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamada, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaide, I., Pesole, G., Quackenbush, J., Schriml, L. M., Staudt, F., Suzuki, P., Tontle, M., Wagner, L., Mashio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bull, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Holtman, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaur, P., Nardone, P., Ring, B., Ringwald, M., Podguz, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weitz, C., Wittaker, C., Wilmink, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohsaki, S., and Hayashizaki, Y.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409 (6821), 685-690 (2001)

5 (bases 1 to 1437)  
Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Aizawa, T., Baldarelli, R., Bono, H., Brownstein, M., Bull, C., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hangaki, T., Hara, A., Hayatsu, N., Hill, D., Hiramoto, K., Hiraoaka, T., Hori, F., Hume, D., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kondo, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Schriml, L., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yamashita, I., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M., and Hayashizaki, Y.  
Direct Submission  
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0645, Japan (E-mail: genome-res@gsr.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)  
Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.

# TITLE JOURNAL COMMENT

CDNA library was prepared and sequenced in Mouse Genome Encyclopedia project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGATCTCGATTAATTAATTAATTCACCCGCCCCCG 3']. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to R0 = 10.0 and subcloning to R0 = 185.2. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATCTCGATTAATTAATTAATTCACCCGCCCCCG 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from lambda FLIC I. Cloning sites, 5' end: SalI, 3' end: BamHI. Host: DH10B.

# FEATURES source

Location/Qualifiers  
1..1437  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="PANTOM:DB:9130017C17"  
/db\_xref="WGC:MGF:1904167"  
/db\_xref="taxon:10090"  
/clone="9130017C17"  
/sex="male"







[illegible]

D6		289	AAGTCCGAAACCGCTGTGGAAAATTTCATGGGAAGAGAATAACAAGCATTTTGACTTTGCCTC	740
OY	424	AGCAAGCACTGACTCTTAATTTCTAATAAT-TAT-ATT-TCT-AAG-GT-CG-AAAG-TCTTG	483	
D6	229	AATAGAACCTGACTCTGAAATTTCTAATAATAATATCTGTAAAGTGTGGGCTGCTTAG	170	
OY	484	CAGAGTAATCTGGGCTGCTTCAAGATCATTCGATGGTGGTGGGCTGCTGGAGTCTGCCG	543	
D6	159	TATATATGACTTGGGCTGCTTCTTAAGAGCATCTGCTGGGCTGGGCTGCTGGAGTCTGG	110	
OY	544	GACTTCTCATCACGCGCGGAGATGAGGAGATTTTCTATTTG-TGT-GAG-TGGACGATAC	603	
D6	109	GACTTCTCATCACGCGCGGAGATGAGGAGATTTTCTATTTG-TGT-GAG-TGGACGATAC	50	
OY	604	CCCAGGCCCCCTGAGCTGCTGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGG	652	
D6	49	CCGAGTCTTGGAGATGATGCTGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGG	1	
RESULT 7				
AM467091/c		506 bp	mRNA linear	EST 24-FEB-2000
LOCUS				
DEFINITION	AM467091	hmo9f03 x1 NC1 CGAP K1d12 Homo sapiens CDNA clone IMAGE:2973213 3'		
ACCESSION	AM467091	mRNA sequence.		
VERSION	AM467091.1	GI:7037197		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Bukacynska, Metazoa, Chordata, Craniata, Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 506) NCI-CGAP http://www.ncbi.nlm.nih.gov/cgi/gap. National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997)			
JOURNAL COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgap@l-trail.rti.gov Tissue Procurement: Christopher Moskajuk, M.D., Ph.D.; Michael R. Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: M. Bento Soares, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LBL at: <a href="#">www.bio.lbl.gov/btsp/image/image.html</a> Seq primer: -40UP from Gibco High quality sequence stop. 472.			
FEATURES SOURCE				
source	location/Qualifiers			
1..506	/organism="Homo sapiens"			
/db_xref="taxon:9606"				
/clone_image="IMAGE.2873213"				
/clone_lib="NCI CGAP Kid12"				
/tissue_type="2 pooled tumors (clear cell type)"				
/lab_host="BDH10B"				
/note="organ: kidney; Vector: pTRITD-Pac (Pharmacia) with a modified polylinker; Site_1: Not 1; Site_2: Eco RI; Plasmid DNA from the normalized library NCI CGAP Kids was prepared, and ss circles were made in vitro. Following BamHI hybridization reaction, this DNA was used as tracer in a subtractive from a pool of 5,000 clones made from the same library (clones: 133912-133831, 147308-147293 and 1492104-149325). Subtraction by Bento Soares and M. Faldut Bonaldo."				
BASE COUNT	96 a 157 C 149 g 104 t			
ORIGIN				
Query Match	51.7%; Score 491.8;	DB 10:	Length 506;	
Best Local Similarity	99.4%;	Pred. No. 8,2e+90;		
Matches 564,	Conservative 0,	Mismatches 2,	Indels 1;	Gaps 1;

OY	146	TGAGCTTCACACTCCAGATTTBACCAAGCGAGCCTTACAGACAGTACAGACAGACAGAAAT	205
Db	506	TGAAGTTCACACTCCAGATTTBACCAAGCGAGCCTTACAGACAGTACAGACAGAAAT	447
OY	206	GATTGTGAAAGTAAGTCAAACCTGCTTTTGGAGTGCCTTGAAGTTCGATTCGAGAAGAA	265
Db	446	GCCCGTGGAAGACGCTCAAACTGCTCTTTGAGATTGCTTCACTCGCATTCGAGAGAGA	387
OY	266	AAGTCTTAAGTTTGTGTGTATCCAAATTCATCGTGCATCTCGAGACCTTTGACAACA	325
Db	386	AAGTCTTAAGTTTGTGTGTATCCAAATTCATCGTGCATCTCGAGACCTTTGACAACA	327
OY	326	ACCAAGCGCTCTCTGGAAACGAGCCTATTCCGACTTCGACGAGTCCGAAAAGCGCTGTGA	385
Db	326	ACCAAGCGCTCTCTGGAAACGAGCCTATTCCGACTTCGACGAGTCCGAAAAGCGCTGTGA	267
OY	386	AGACGTTCAAGSAGAGAGATCGAAGACGTGTGAGCTTCCCAGAGAACCCTGACTGSGAACT	445
Db	266	AGACGTTCAAGSAGAGAGATCGAAGACGTGTGAGCTTCCCAGAGAACCCTGACTGSGAACT	207
OY	446	TGCTGAGAGAGATGATTCGTGTGAGCCTCGGCGGCCCTGAGAGAGTACCTGGGCGCTCT	505
Db	206	TGCTGAGAGAGATGATTCGTGTGAGCCTCGGCGGCCCTGAGAGAGTACCTGGGCGCTCT	147
OY	506	ACGCCATCCGCTGCTGCGCGCCGCTCCCGGAGATTCCTGTGACTTCACAGCGCGCGAGC	565
Db	146	ACGCCATCCGCTGCTGCGCGCCGCTCCCGGAGATTCCTGTGACTTCCTCACGCGCGGAGC	87
OY	566	TGCGGAGAGCTTTTGTGTGCTGCTGCGCGCGCTGCGAGTACCTGTGAGCTGTGCTGC	625
Db	86	TGCGGAGAGCTTTTGTGTGCTGCTGCGCGCGCTGCGAGTACCTGTGAGCTGTGCTGC	27
OY	626	TGCGGAGAGCTGCGCGCGCTGCGAGAGAGC	652
Db	26	TGCGGAGAGCTGCGCGCGCTGCGAGAGAGC	1
<hr/>			
RESULT 8			
LOCUS	A1004982/c		
DEFINITION	AI004982 478 bp mRNA linear EST 27-AUG-1998		
IMAGE	00072409.x1 Soares_NEL_T_GBC_S1 Homo sapiens cDNA clone IMAGE1632560 3', mRNA sequence.		
ACCESSION	A1004982		
VERSION	A1004982.1		GI:3214492
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
JOURNAL	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 (bases 1 to 478)		
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/hic/gap.		
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index		
COMMENT	Unpublished (1997) Contact: Robert Strausberg, Ph.D. Email: cgapsb@mail.nih.gov This clone is available royalty-free through LNL; contact the IMAGE Consortium (infoimage.lnl.gov) for further information. Insert Length: 791 Std Error: 0.00 Seq primer: primer name ambiguous.		
FEATURES	Location/Qualifiers		
source	1..478		

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1625560"
/clone_1ib="Scatter_NFL_T_GBC_S1"
/lab_host="DH10B"
/notes="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not I; Site 2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NHHL3W, testis NHT, and B-cell
NCI-GC98-GCB1) were mixed, and ss circles were made in

```

vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of 1.M.A.G.E. clones 227480-102087, 622622-667792, 726408-722871, and 722696-733392, subtraction by Bento Soares and M. Fatima Bonaldo. "

Query Match	49.4%	Score 470;	DB 9;	Length 478;
Best Local Similarity	99.0%;	Pred. No. 2.1e-85;		
Matches 473;	Conservative 0;	Mismatches 5;	Indels 0;	Gaps 0;

QY	200	AGAAATGCGCTGSAAGACACGTCAACAGCTCTTTGAGATACCTTACAGCTGCATCGAGG	259
Db	478	AGAAATGCGCGTGSAAAGACACGTCCACAGCTCTTTGAGATATGCTTTACATGATCGAGG	419
QY	260	AGAGAAAAGTCTCTAAAGTTTGTGTGTATCAAAATCATGCTGATCCAGATTTGAGAGCTTTG	319
Db	418	AGAGAAAAGTCTCTAAAGTTTGTGTGTATCAAAATCATGCTGATCCAGATTTGAGAGCTTTG	359
QY	320	ACACACAACAAGACCGTCTCTGSAAGACCGCTATTTCTGCAATTCCTGSAAGATTCACAAAGACCG	379
Db	358	ACACACAACAAGACCGTCTCTGSAAGACCGCTATTTCTGCAATTCCTGSAAGATTCACAAAGACCG	299
QY	380	TGCTGAGAGACGTTTCAGGAGAGATGGAAGACGTTGAGATTCTCCAGAGACACCTTACATG	439
Db	298	TGCTGAGAGACGTTTCAGGAGAGATGGAAGACGTTGAGATTCTCCAGAGACACCTTACATG	239
QY	440	GGAACCTTCGCTGAGAGAGATGATCTTGTAGACGCTGAGACGCTTGTAGAGATATCTGAGCTC	499
Db	238	GGAACCTTCGCTGAGAGAGATGATCTTGTAGACGCTGAGACGCTTGTAGAGATATCTGAGCTC	179
QY	500	TGACTTACGCCATCTGCTGCTGAGCGCGCTCTCCGAGAGTCTCTGACCTTCTCAGCGCGC	559
Db	178	TGCTCTACGCCATCTCGCTGCTGAGCGCGCTCTCCGAGAGTCTCTGACCTTCTCAGCGCGC	119
QY	560	CGAGCTCGCGCAAGCTTTCGATCTGAGAGCTGAGAGACCTCAACGCAATGCTGCTGCGCGCGC	619
Db	118	CGAGAGCTGAGAGAGCTTTCGATCTGAGAGCTGAGAGACCTCAACGCAATGCTGCTGCGCGCGC	59
QY	620	TGCTGCTGAGCGCTGCTGCTGCTGAGAGAGACCTCAACGCAATGCTGCTGCGCGCGC	677
Db	58	TGCTGCTGAGCGCTGCTGCTGCTGAGAGAGACCTCAACGCAATGCTGCTGCGCGCGC	1

RESULT 9	BI099756	671 bp	mRNA	linear	EST 26 JUN 2001
LOCUS	BI099756				
DEFINITION	BI099756	602884941F1	NCI_CGAP_Kid1a Mus musculus	CDNA clone	IMAGE:5040274
ACCESSION	BI099756	5', mRNA sequence,			
VERSION	BI099756.1	GI:14550649			
KEYWORDS	EST.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 671)				
AUTHORS	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .				
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)				
JOURNAL	unpublished (1999)				
COMMENT	Contact: Robert Strausberg, Ph.D.				

Email: cgaabds-r@mail.nih.gov  
 Tissue Procurement: Jeffrey E. Green, M.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MSC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLNL1111 row: a column: 11



CY	659	TGTCCTGGCGTGTATGCTGTTGAAGAAAGAAGTGACGCGC	659
DB	40	TGTTCTGTGTAATTGCTGCTGCTGCAGDGAAGACTCACCCGC	1
RESULT	11		
Accession	AM592491	460 bp	mRNA linear EST 22-MAR-2000
Definition	h43c11.x1 Scapex NFL T GGC_S1 Homo sapiens cDNA clone IMAGE:2934644 3' mRNA sequence.		
Accession	AM592491		
Version	AM592491.1	GI:7279674	
Keywords	EST.		
Source	human.		
Organism	Homo sapiens		
Euarchyotax:	Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
Mammalia;	Eutheria; Primates; Catarrhini; Homnidae; Homo.		
Reference	1 (bases 1 to 460)		
Authors	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.		
Title	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index		
Journal	Unpublished (1997)		
Comment	Contact: Robert Strausberg, Ph.D. Email: cgaps-tr@mail.nih.gov This clone is available royalty-free through LNL ; contact the IMGE Consortium (info@image.jnl.gov) for further information. Possible reversed clone: polyt not found Seq primer: -40UP from Gibco High quality sequence stop: 459.		
Features	location/Qualifiers		
Source	1..460		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:2934644"		
	/clone_lib="Soares_NFL_T_GBC_SI"		
	/lab_host="NDH10B"		
	/note="organ: pooled; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NDH19W, testis NH7, and B-cell NCI CGAP GCBI) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of 1 M A.G.E. clones 297480-302687, 682632 687239, 726008-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo."		
BASE COUNT	87 a 145 c 136 g 92 t		
ORIGIN			
Query Match	48.4%; Score 460; DB 10; Length 460;		
Best Local Similarity	100.0%; Fred.No. 2.2e-83;		
Matches 460; Conservative	0; Mismatches 0; Indels 0; Gaps 0;		
CY	131	CAGAACCCAGAAATGCTGCTGTAAGTTCGTGTACCAATCATGCTATCCGAACTGGG	252
DB	460	CAGAACCCAGAAATGCCGCTCGAAGACGCGCAACCTGCTTTGAGATCGCTTAGCTGCG	401
CY	253	ATCGAGAGAGAAAAAGTCTCTAAGTTTGTGTGTACCAATCATGCTATCCGAACTGGG	312
DB	400	ATCGAGAGAGAAAAAGTCTCTAAGTTTGTGTGTACCAATCATGCTATCCGAACTGGG	341
CY	313	ACCTTTGAAACAACAAGGCGCTCTCTGGAAGGCGGTATTCGACTTCGGGAAGCTCCAG	372
DB	340	ACCTTTGAAACAACAAGGCGCTCTCTGGAAGGCGGTATTCGACTTCGGGAAGCTCCAG	281
CY	373	AAAAGCGCTTCTGAAGAAGTCTTGAAGTAAGAGAGCAAGAGTGTAGTTTCCAGAAAGC	432
DB	280	AAAAGCGCTTCTGAAGAAGTCTTGAAGTAAGAGAGCAAGAGTGTAGTTTCCAGAAAGC	221
CY	433	CTGACTGTGTAATCTTGTGTGAAGAGATGATCTTGAAGTGTGCGTGTGAGAGATG	492

[illegible]

RESULT 13	
A1131560/c	
LOCUS	459 bp mRNA linear EST 14-SEP-1998
DEFINITION	CB882804.xl Soares_fetal_heart_Nbhhl9w Homo sapiens cdna clone
	IMAGE:1706599 3', mRNA sequence.
ACCESSION	A1131560
VERSION	A1131560.1 GI:3601576

FEATURES	Location/Qualifiers
SOURCE	1. .459

```

/organism="Homo sapiens"
/db xref="taxon:9606"
/clone="IMAGE:1706599"
/clone_lib="Soares_fetal_heart_NBHL19W"
/sex="unknown"
/dev stage="19 weeks"
/lab host="NH108 (ampicillin resistant)"
/notes="Organ: heart; Vector: pT73d (Pharmacia) with a
modified polylinker. Site 1: Not 1; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I...ctggc(gt) primer [5'
TGTTCACATCTCTGAAGTGGAGCGCGCCATCTTTTCTTTTCTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M.Patricia Bonaldi. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NBHL19W."

```

Query Match	48.3%	Score 459;	DB 9;	Length 459;
Best Local Similarity	100.0%	Pred. No. 3,66	83,	
Matches 459; Conservative	0;	Mismatches	0;	Gaps 0;

[illegible]

RESULT 14					
AM469821/c					
LOCUS	457 bp	mRNA	linear	EST 24 FEB 2000	
DEFINITION	hd3f5f08_xl Soares NFL_T_GGC-St Homo sapiens cDNA clone				
IMAGE:	2211527 3'	mRNA sequence.			

ACCESSION	AM469821
VERSION	AM469821.1
	GI.7039927

KEYWORDS	EST.
SOURCE	human.

ORGANISM Homo sapiens  
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE  
1. (pages 1 to 459)  
AUTHORS  
NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL, unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-rc@mail.nih.gov](mailto:cgapbs-rc@mail.nih.gov)  
This clone is available royalty free through TINT. Contact the

IMAGE Consortium ([infoimage.ihl.gov](http://infoimage.ihl.gov)) for further information.

possible reversed clone: polyt not found  
Seq primer: -40UP from Gibco

High quality sequence stop: 408.  
Location/Qualifiers

SOURCE

1. 459

/organism- "Homo canianci"

```

/organism="Homo sapiens"
/db_xref="taxon:9606"

```

```
/clone="IMAGE:2911527"  
/clone_lib="Soares_NFL_T_GBC_S1"
```

```
/lab_host="DH10B"
/notef="Orgar: pooled: Vector: pT7T3D-Pac (Pharmacia) with
```

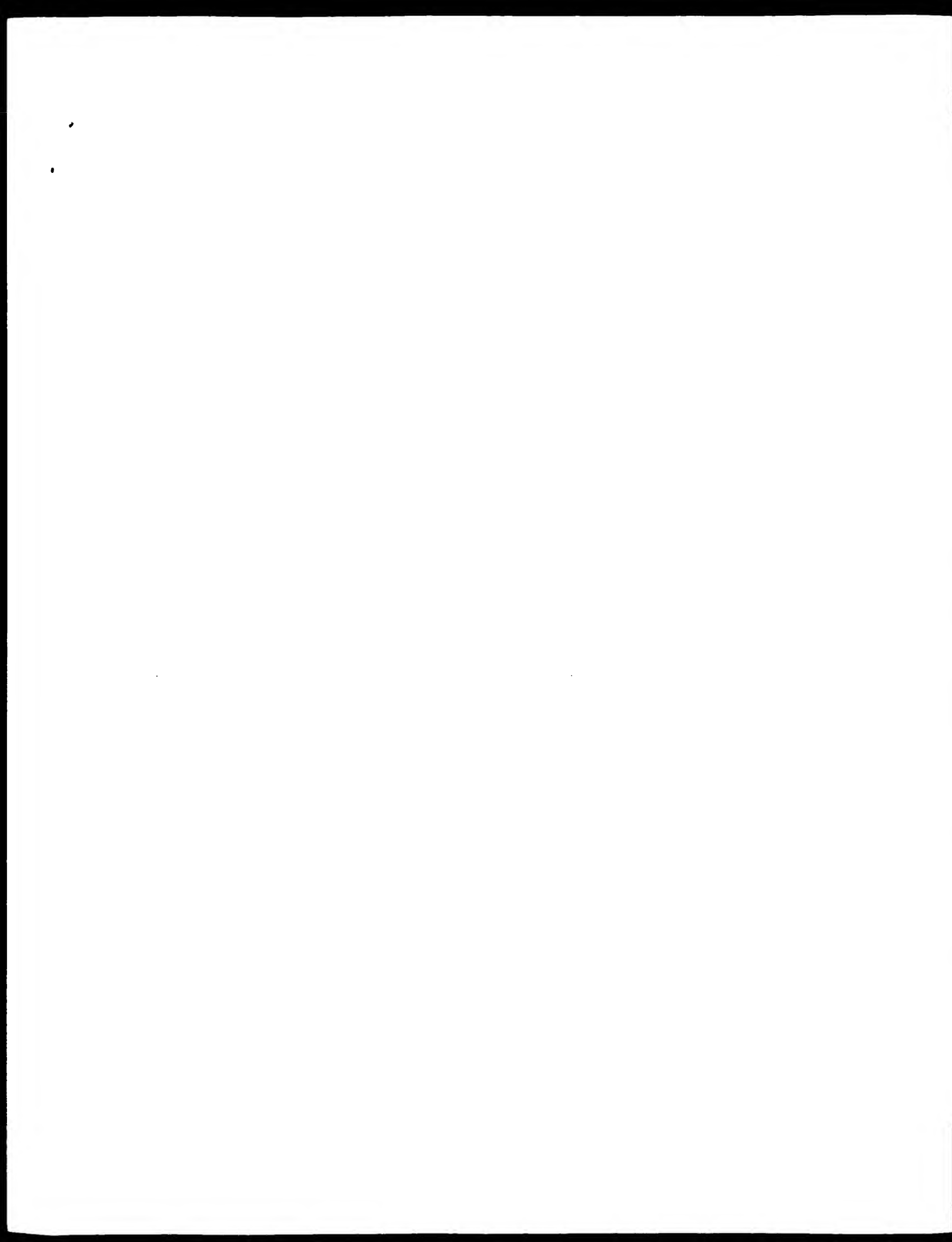
a modified polylinker; Site\_1: Not I, Site\_2: Eco RI;

Equal amounts of plasmid DNA from three normalized libraries (fetal lung NBHL19W, testis NHT, and B-cell NCI-GAP GCBI) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of 1 M A G E clones 297480-302987, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bernaldo.

BASE COUNT 88 a 145 c 133 g 93 t  
 ORIGIN  
 Query Match 47.9%; Score 455.8; DB 10; Length 459;  
 Best Local Similarity 90.6%; Pred. No. 1.66-92;  
 Matches 457; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 154 AAGAACTAGAAATGCTGCTGAGTGTGTTGTTGCAATATCATCTCATCCAGACTGGGA 313  
 Db 459 AAGAACTAGAAATGCTGCTGAGTGTGTTGTTGCAATATCATCTCATCCAGACTGGGA 400  
 QY 254 TCGAG 313  
 Db 309 TCGAG 340  
 QY 314 GCTTGAACATACAG 373  
 Db 339 GCTTGAACATACAG 280  
 QY 374 AAG 433  
 Db 279 AAG 220  
 QY 434 TGAATGAG 493  
 Db 219 TGAATGAG 160  
 QY 434 TGAATGAG 553  
 Db 159 TGAATGAG 200  
 QY 434 TGAATGAG 613  
 Db 323 TGAATGAG 40  
 QY 614 TGAATGAG 652  
 Db 323 TGAATGAG 1  
 RESULT 15  
 A1554174/c 500 bp mRNA linear EST 13-APR-1999  
 LOCUS t60605.x1 Soares\_NFL\_T\_GBC\_S1 Homo sapiens cDNA clone  
 DEFINITION IMAGE:2090121.3, mRNA sequence.  
 ACCESSION A1554174  
 VERSION A1554174.1 GI:4486537  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE NCI-GAP http://www.ncbi.nlm.nih.gov/ncigap.  
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 TITLE Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: c9aps-remail.nih.gov  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (infoimage.lnl.gov) for further information.  
 Insert Length: 838 Std Error: 0.00  
 Seq primer: -40UP from GIBCO

FEATURES  
 source  
 High quality sequence stop: 491.  
 Location/Qualifiers  
 1..500  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2090121"  
 /cdate="11b=Soares\_NFL\_T\_GBC\_S1"  
 /lab="lab=VH10B"  
 /note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with  
 a modified polylinker, site 1: Not 1; Site 2: Eco RI;  
 Equal amounts of plasmid DNA from three normalized  
 libraries (fetal lung NBHL19W, testis NHT, and B-cell  
 NCI-GAP GCBI) were mixed, and ss circles were made in  
 vitro. Following HAP purification, this DNA was used as  
 tracer in a subtractive hybridization reaction. The driver  
 was PCR-amplified cDNAs from pools of 5,000 clones made  
 from the same 3 libraries. The pools consisted of  
 1 M A G E clones 297480-302987, 682632-687239,  
 726408-728711, and 729096-731399. Subtraction by Bento  
 Soares and M. Fatima Bernaldo."

BASE COUNT 94 a 155 c 154 g 96 t 1 others  
 ORIGIN  
 Query Match 47.7%; Score 454; DB 9; Length 500;  
 Best Local Similarity 99.0%; Pred. No. 3.86-82;  
 Matches 498; Conservative 0; Mismatches 1; Indels 4; Gaps 4;  
 QY 176 AGCTCAG 235  
 Db 500 AGCTCAG 443  
 QY 236 AGATGCG-ITCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 294  
 Db 442 AGATGCGITTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 383  
 QY 225 ATGCTCAG 354  
 Db 382 ATGCTCAG 324  
 QY 355 GAGCTCAG 414  
 Db 323 GAGCTCAG 264  
 QY 415 GAGCTCAG 474  
 Db 263 GAGCTCAG 204  
 QY 475 GCGGCGCTGAG 534  
 Db 203 GCGGCGCTGAG 144  
 QY 535 GAGCTCAG 594  
 Db 143 GAGCTCAG 84  
 QY 595 GAGCTCAG 654  
 Db 83 GAGCTCAG 24  
 QY 655 ACCGCGCTGAG 677  
 Db 23 ACCGCGCTGAG 1  
 Search completed: April 20, 2003, 16:58:31  
 Job time: 3723 secs









bowel disease 1 (IBD1) protein which is associated with intestinal inflammatory disease. The IBD1 gene is probably involved in regulation of apoptosis and activation of NF-kappa B. The IBD1 and IBD1prox polymorphotides are as useful as source of probes and primers, as source of (anti)sense oligonucleotides, for recombinant production of polypeptides, and in screening for interactive compounds. The polypeptides are used to raise specific antibodies which useful for diagnostic detection or purification of IBD1 and IBD1prox, to screen for specific binding agents, potential therapeutic agents. The IBD1 and IBD1prox polymorphotides and polypeptides are useful for treatment and prevention of inflammatory and/or immune diseases or cancer, where associated with mutations in genes corresponding to IBD1 and IBD1prox, especially cryptogenic inflammation of the intestines (hemorrhagic rectocolitis, Crohn's disease and Blau syndrome).

Sequence 1315 BP; 257 A, 419 C, 394 G, 245 T, 3 other

Query Match	99.8%	Score 949.4	DB 22	Length 1315
Best Local Similarity	99.98%	Pred. No. 2.4e-166		
Matches 950; Conservative	0	Mismatches 1	Indels 0	Gaps 0

[illegible]







XX Human; cytosine methylation; 5'-CpG 3', uracil; cytosine; diagnosis;  
 KM drug; side effect; cancer; central nervous system; cardiovascular;  
 KM gastrointestinal; respiratory system; single nucleotide polymorphism;  
 KM SNP; cell differentiation; ds.  
 XX  
 OS Homo sapiens.  
 FN WC000218632-A2.  
 XX  
 PD 07-MAR-2002.  
 XX  
 PE 01-SEP-2001; 2001WO-EP10074.  
 XX  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX  
 PR 05-SEP-2000; 2000DE-1044543.  
 XX  
 PA (EPIC-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
 XX  
 DR WPI; 2002-371829/40.  
 XX  
 PT Determining the degree of cytosine methylation in genomic DNA, useful  
 PT for diagnosis and prognosis, comprises selective hybridization of  
 PT amplicons from chemically treated DNA -  
 PS  
 PS Claim 12; 56pp + Sequence Listing; 56pp; German.  
 XX  
 CC This invention describes a novel method for determining the degree of  
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
 CC genomic sample of DNA. The sample is treated chemically to convert  
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
 CC DNA that contains the target C is amplified to form a labeled amplicon.  
 CC The amplicon is hybridised to two classes, each with at least one  
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
 CC and the degree of hybridisation to both classes is determined from the  
 CC label on the amplicon. From the ratio of labels hybridised to the two  
 CC classes of oligomers, the degree of methylation is calculated. The method  
 CC is used: (i) for diagnosis and/or prognosis of side effects of  
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
 CC systems etc., particularly by detecting mutations or single nucleotide  
 CC polymorphisms (SNPs); and (ii) for differentiation of cell or tissue  
 CC types and for investigating cell differentiation. The method allows the  
 CC methylation status of many C residues to be determined simultaneously.  
 CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the  
 CC method for determining the degree of cytosine methylation described in  
 CC the disclosure of the invention.  
 CC  
 XX  
 XX Sequence 570 BP; 216 A; 203 C; 62 G; 90 T, 0 other;  
 SO  
 Query Match 37.2%; Score 354; DB 24; Length 570;  
 Best Local Similarity 76.1%; Pred No 14e-56;  
 Matches 435; Conservative 0; Mismatches 135; Indels 0; Gaps 0;  
 Yr 343 CGGGCTATTCGATTTCTGAGTTCGAGAAAGGCTGCTGAGAGCTTACGAGAGAG 402  
 Db 1 CGAGCTATTCGAGTTCGAGAAAGGCTGAGAAAGGCTGAGAGCTTACGAGAGAG 60  
 Yr 403 ATGGAAGATGAGATTCGAGTTCGAGAAAGGCTGAGAAAGGCTGAGAGATGATG 462  
 Db 61 ATGGAAGATGAGATTCGAGTTCGAGAAAGGCTGAGAAAGGCTGAGAGATGATG 120  
 Yr 463 TATGAGAGATGAGATTCGAGTTCGAGAAAGGCTGAGAAAGGCTGAGAGATGATG 522  
 Db 121 TATGAGAGATGAGATTCGAGTTCGAGAAAGGCTGAGAAAGGCTGAGAGATGATG 180  
 Yr 523 CGGGCTATTCGAGTTCGAGTTCGAGAAAGGCTGAGAAAGGCTGAGAGATGATG 582  
 Db 181 CGGGCTATTCGAGTTCGAGTTCGAGAAAGGCTGAGAAAGGCTGAGAGATGATG 240  
 Yr 583 TGGCTATTCGAGTTCGAGTTCGAGAAAGGCTGAGAAAGGCTGAGAGATGATG 642

Db 241 TACCTACGAAACCGAACCAATACCCGCGCGCTTAAATATTAATGATGATGCTA 300  
 Yr 643 CAGAGAGAGATGAGATTCGAGTTCGAGAAAGGCTGAGAAAGGCTGAGAGATGATG 702  
 Db 301 CAAAAAAAGTACGAGTTCGAGTTCGAGAAAGGCTGAGAAAGGCTGAGAGATGATG 360  
 Yr 703 CTGTGCGACCGGAGATTCGAGTTCGAGAAAGGCTGAGAAAGGCTGAGAGATGATG 762  
 Db 361 CTATACGACCGGAGATTCGAGTTCGAGAAAGGCTGAGAAAGGCTGAGAGATGATG 420  
 Yr 763 CAGCGCTGAGAGATTCGAGTTCGAGAAAGGCTGAGAAAGGCTGAGAGATGATG 822  
 Db 421 CAAAGGCTGAGAGATTCGAGTTCGAGAAAGGCTGAGAAAGGCTGAGAGATGATG 480  
 Yr 823 CGCTGCGCTGAGAGATTCGAGTTCGAGAAAGGCTGAGAAAGGCTGAGAGATGATG 882  
 Db 481 CGCTGAGATTCGAGTTCGAGTTCGAGAAAGGCTGAGAAAGGCTGAGAGATGATG 540  
 Yr 883 CAGCTCGAGAGATTCGAGTTCGAGAAAGGCTGAGAAAGGCTGAGAGATGATG 912  
 Db 541 CAACTCGAGAGATTCGAGTTCGAGAAAGGCTGAGAAAGGCTGAGAGATGATG 570  
 RESULT 8  
 ABQ18058  
 ID ABQ18058 standard; DNA; 570 BP.  
 XX  
 AC ABQ18058;  
 XX  
 DT 12-JUL-2002 (first entry)  
 XX  
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 4649.  
 XX  
 KW Human; cytosine methylation; 5'-CpG-3', uracil; cytosine; diagnosis;  
 KW drug; side effect; cancer; central nervous system; cardiovascular;  
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
 KW SNP; cell differentiation; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 FN WC000218632-A2.  
 XX  
 PD 07-MAR-2002.  
 XX  
 PE 01-SEP-2001; 2001WO-EP10074.  
 XX  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX  
 PR 05-SEP-2000; 2000DE-1044543.  
 XX  
 PA (EPIC-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
 XX  
 DR WPI; 2002-371829/40.  
 XX  
 PT Determining the degree of cytosine methylation in genomic DNA, useful  
 PT for diagnosis and prognosis, comprises selective hybridization of  
 PT amplicons from chemically treated DNA -  
 PS  
 PS Claim 12; 56pp + Sequence Listing; 56pp; German.  
 XX  
 CC This invention describes a novel method for determining the degree of  
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
 CC genomic sample of DNA. The sample is treated chemically to convert  
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
 CC DNA that contains the target C is amplified to form a labeled amplicon.  
 CC The amplicon is hybridised to two classes, each with at least one  
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
 CC and the degree of hybridisation to both classes is determined from the  
 CC label on the amplicon. From the ratio of labels hybridised to the two  
 CC classes of oligomers, the degree of methylation is calculated. The method  
 CC is used: (i) for diagnosis and/or prognosis of side effects of







[illegible]







PR 06-SEP-2000; 2000US-0230437  
 PR 06-SEP-2000; 2000US-0230438  
 PR 06-SEP-2000; 2000US-0231242  
 PR 06-SEP-2000; 2000US-0231243  
 PR 06-SEP-2000; 2000US-0231244  
 PR 06-SEP-2000; 2000US-0231413  
 PR 06-SEP-2000; 2000US-0231414  
 PR 06-SEP-2000; 2000US-0232060  
 PR 06-SEP-2000; 2000US-0232081  
 PR 12-SEP-2000; 2000US-0231968  
 PR 14-SEP-2000; 2000US-0232397  
 PR 14-SEP-2000; 2000US-0232398  
 PR 14-SEP-2000; 2000US-0232399  
 PR 14-SEP-2000; 2000US-0232400  
 PR 14-SEP-2000; 2000US-0232401  
 PR 14-SEP-2000; 2000US-0233063  
 PR 14-SEP-2000; 2000US-0233064  
 PR 21-SEP-2000; 2000US-0233065  
 PR 21-SEP-2000; 2000US-0234223  
 PR 25-SEP-2000; 2000US-0234997  
 PR 25-SEP-2000; 2000US-0234998  
 PR 26-SEP-2000; 2000US-0235484  
 PR 27-SEP-2000; 2000US-0235834  
 PR 27-SEP-2000; 2000US-0235836  
 PR 29-SEP-2000; 2000US-0236327  
 PR 29-SEP-2000; 2000US-0236327  
 PR 29-SEP-2000; 2000US-0236367  
 PR 29-SEP-2000; 2000US-0236368  
 PR 29-SEP-2000; 2000US-0236369  
 PR 29-SEP-2000; 2000US-0236370  
 PR 02-OCT-2000; 2000US-0236802  
 PR 02-OCT-2000; 2000US-0237037  
 PR 02-OCT-2000; 2000US-0237038  
 PR 02-OCT-2000; 2000US-0237039  
 PR 02-OCT-2000; 2000US-0237040  
 PR 13-OCT-2000; 2000US-0239935  
 PR 13-OCT-2000; 2000US-0239937  
 PR 20-OCT-2000; 2000US-0240960  
 PR 20-OCT-2000; 2000US-0241221  
 PR 20-OCT-2000; 2000US-0241285  
 PR 20-OCT-2000; 2000US-0241786  
 PR 20-OCT-2000; 2000US-0241787  
 PR 20-OCT-2000; 2000US-0241808  
 PR 20-OCT-2000; 2000US-0241809  
 PR 01-NOV-2000; 2000US-0244617  
 PR 08-NOV-2000; 2000US-0246474  
 PR 08-NOV-2000; 2000US-0246475  
 PR 08-NOV-2000; 2000US-0246476  
 PR 08-NOV-2000; 2000US-0246477  
 PR 08-NOV-2000; 2000US-0246478  
 PR 08-NOV-2000; 2000US-0246523  
 PR 08-NOV-2000; 2000US-0246524  
 PR 08-NOV-2000; 2000US-0246525  
 PR 08-NOV-2000; 2000US-0246526  
 PR 08-NOV-2000; 2000US-0246527  
 PR 08-NOV-2000; 2000US-0246528  
 PR 08-NOV-2000; 2000US-0246529  
 PR 08-NOV-2000; 2000US-0246530  
 PR 08-NOV-2000; 2000US-0246610  
 PR 08-NOV-2000; 2000US-0246611  
 PR 17-NOV-2000; 2000US-0249207  
 PR 17-NOV-2000; 2000US-0249208  
 PR 17-NOV-2000; 2000US-0249209  
 PR 17-NOV-2000; 2000US-0249210  
 PR 17-NOV-2000; 2000US-0249211  
 PR 17-NOV-2000; 2000US-0249212  
 PR 17-NOV-2000; 2000US-0249213  
 PR 17-NOV-2000; 2000US-0249214  
 PR 17-NOV-2000; 2000US-0249215  
 PR 17-NOV-2000; 2000US-0249216  
 PR 17-NOV-2000; 2000US-0249217

PR 17-NOV-2000; 2000US-0249218  
 PR 17-NOV-2000; 2000US-0249244  
 PR 17-NOV-2000; 2000US-0249245  
 PR 17-NOV-2000; 2000US-0249246  
 PR 17-NOV-2000; 2000US-0249265  
 PR 17-NOV-2000; 2000US-0249267  
 PR 17-NOV-2000; 2000US-0249269  
 PR 17-NOV-2000; 2000US-0249300  
 PR 01-DEC-2000; 2000US-0250160  
 PR 01-DEC-2000; 2000US-0250397  
 PR 05-DEC-2000; 2000US-0251030  
 PR 05-DEC-2000; 2000US-0251988  
 PR 06-DEC-2000; 2000US-0256719  
 PR 06-DEC-2000; 2000US-0251479  
 PR 08-DEC-2000; 2000US-0251856  
 PR 08-DEC-2000; 2000US-0251868  
 PR 08-DEC-2000; 2000US-0251869  
 PR 08-DEC-2000; 2000US-0251889  
 PR 08-DEC-2000; 2000US-0251989  
 PR 11-DEC-2000; 2000US-0251990  
 PR 05-JAN-2001; 2001US-0259678  
 PR  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Barash SC, Ruben SM;  
 XX WPI, 2001-476224/51.  
 XX  
 PT Isolated polypeptide for treating, preventing and/or prognosing  
 PT disorders related to the respiratory system including respiratory  
 FT cancers and also for testing and detection e.g. diagnosis  
 XX  
 PS Disclosure; SED ID No 990; 546pp; English.  
 CC  
 CC The present invention relates to the isolation of novel human  
 CC respiratory antigens (AAU17685-AAU17975), and cDNA and genomic  
 CC sequences encoding for these polypeptides. The sequences of the  
 CC invention are useful for preventing, treating and/or prognosing  
 CC disorders related to the respiratory system including throat  
 CC disorders (e.g. vocal cord paralysis, tonsillitis, and laryngitis),  
 CC lung disorders e.g. pneumonia, allergic disorders e.g. asthma,  
 CC pleurisy, cystic fibrosis, emphysema, nose disorders and cancers of  
 CC the respiratory tissues e.g. lung cancer. The polynucleotide sequences  
 CC of the invention are useful in gene therapy and antisense therapy.  
 CC AA32861-AA32874 represent genomic sequences encoding for novel  
 CC human respiratory antigens.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pcl\_sequences.  
 XX  
 SO Sequence 23586 BP, 5938 A, 5961 C, 5927 G, 5754 T, 0 other;  
 Query Match 14.3%, Score 135.8, DB 22; Length 23580;  
 Best Local Similarity 53.6%; Pred. No. 2.9e-16;  
 Matches 306; Conservative 0; Mismatches 266; Indels 3; Gaps 1;  
 QY 282 GGTGTACCAATATGATGATATCCAGATGAGCTTGAACACACAGAGCGGCTCTGGA 341  
 DB 6828 GCTTACACCTGCTGCTGATATGCTGAGACCGTCAATTTCCGACCAATCTC 6887  
 QY 342 ATGGGCTATTTGGAATTTGAGATTTGAGAAAGGCTGCTGAGAGCTTGAAGAGGA 401  
 DB 6988 TGCCCTTACTGAGATTTGAGAGATTTGAGAGAACTTGAAGATTTGAGAGAG 4947  
 QY 402 GATCGAAGAGCTGAGATTTCCAGAGAGTAATTAATGAGAACTTGAAGAGATGAT 461  
 DB 6948 AATGCTGCTGATCTCTCTTCTGCTGAGAGAGCTGAGAGAACTTGAAGAGAT 7007  
 QY 462 CTGTAGAGCTGAGAGCTGCTGAGAGATTTGAGAGATTTGAGAGATTTGAGAGAT 521  
 DB 7008 TATGCTGATGAGAGAGCTTGAAGAGATTTGAGAGATTTGAGAGAGATTTGAG 7067  
 QY 522 GCGGCTGCTGAGAGCTTGAAGAGATTTGAGAGAGCTGAGAGAGATTTGAGAG 581



[illegible]

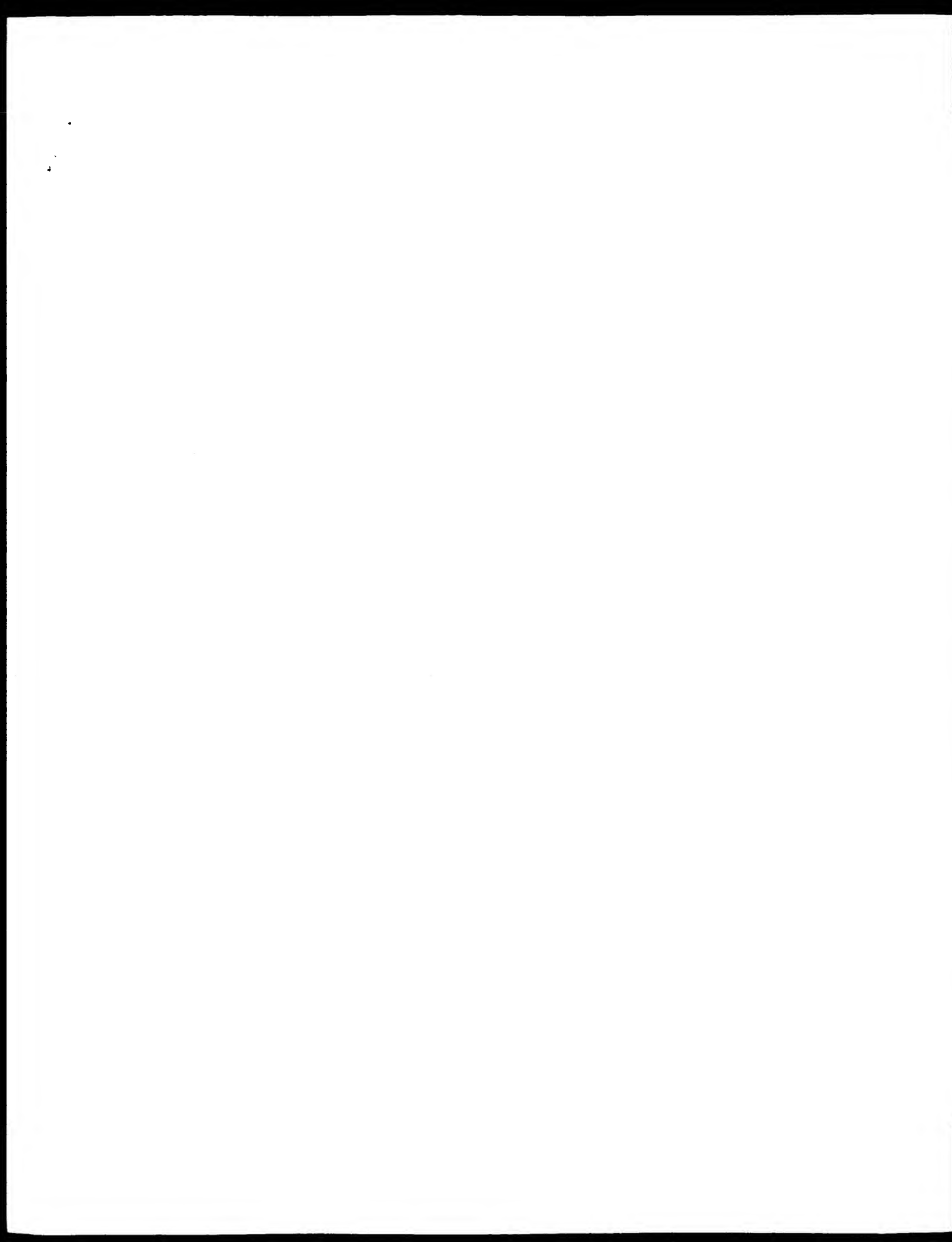
```

Query Match      14.3%   Score 135.8; DB 23; Length 23580;
Best Local Similarity 53.6%; Pident No. 2,9e+16;
Matches 306; Conservative 0; Mismatches 262; Indels 3; Gaps 1

```

[illegible]

Search completed. April 20, 2003, 15:57:11  
Job time : 975 secs







```

Db 181 CAGACTACTGCGAGAACGAAATGCGCTGGAGACAGCTCAAACTGCTTTAGATC 240
QY 241 GCTTCAGCTTGGCATGAGAGAGAAAAGTCTTAAGTTTGTGTGTACCAATCATCTC 300
Db 241 GCTTCAGCTTGGCATGAGAGAGAAAAGTCTTAAGTTTGTGTGTACCAATCATCTC 300
QY 301 ATCCAGACTGAGAGCTTTTACACACACAGAGGCTGCTGGAAAGGCTATCTCCAGCTC 360
Db 301 ATCCAGACTGAGAGCTTTTACACACACAGAGGCTGCTGGAAAGGCTATCTCCAGCTC 360
QY 361 GGTAAATCTTCAAAATCTTCAAAATCTTCAAAATCTTCAAAATCTTCAAAATCTT 420
Db 361 GGTAAATCTTCAAAATCTTCAAAATCTTCAAAATCTTCAAAATCTTCAAAATCTT 420
QY 421 CCGAGAGAGAGCTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTT 480
Db 421 CCGAGAGAGAGCTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTT 480
QY 481 CTAATTAATTAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTT 540
Db 481 CTAATTAATTAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTT 540
QY 541 CTGCAATTTTCTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCA 600
Db 541 CTGCAATTTTCTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCA 600
QY 601 TACCGAGAGAGCTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTT 660
Db 601 TACCGAGAGAGCTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTT 660
QY 661 CATTTTCTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTT 720
Db 661 CATTTTCTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTT 720
QY 721 GACCGAGAGAGCTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTT 780
Db 721 GACCGAGAGAGCTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTT 780
QY 781 GAGAGAGAGAGCTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTT 840
Db 781 GAGAGAGAGAGCTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTT 840
QY 841 GGTAAATCTTCAAAATCTTCAAAATCTTCAAAATCTTCAAAATCTTCAAAATCTT 900
Db 841 GGTAAATCTTCAAAATCTTCAAAATCTTCAAAATCTTCAAAATCTTCAAAATCTT 900
QY 901 CCGAGAGAGAGCTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTT 951
Db 901 CCGAGAGAGAGCTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTT 951

```

## RESULT 2

```

US-09-764-860-990
; Sequence 990, Application US/09764860
; Patent No. US2002009493A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC008
; CURRENT APPLICATION NUMBER: US/09/764,860
; PRIORITY FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1198
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 990
; LENGTH: 23580
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-860-990

```

```

Query Match 14.3%; Score 135.8; DB 10; Length 23580;
Best Local Similarity 53.6%; Pred. No. 2,5e-25;
Matches 306; Conservative 0; Mismatches 262; Indels 3; Gaps 1;

```

```

QY 282 GGTGTAACCAATCATGATCATCAATTTTCAAAATTTTCAAAATTTTCAAAATTTT 341
Db 6828 GGTGTAACCAATCATGATCATCAATTTTCAAAATTTTCAAAATTTTCAAAATTTT 341
QY 342 AGCGAGAGAGCTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTT 401
Db 6888 AGCGAGAGAGCTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTT 401
QY 402 GATGAGAGAGCTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTT 461
Db 6948 GATGAGAGAGCTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTT 461
QY 462 CTGCAATTTTCTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTT 521
Db 7008 CTGCAATTTTCTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTT 521
QY 522 GCGAGAGAGAGCTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTT 581
Db 7068 GCGAGAGAGAGCTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTT 581
QY 582 CTGCAATTTTCTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTT 641
Db 7128 CTGCAATTTTCTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTT 641
QY 642 GGTAAATCTTCAAAATCTTCAAAATCTTCAAAATCTTCAAAATCTTCAAAATCTT 701
Db 7188 GGTAAATCTTCAAAATCTTCAAAATCTTCAAAATCTTCAAAATCTTCAAAATCTT 701
QY 701 GGTAAATCTTCAAAATCTTCAAAATCTTCAAAATCTTCAAAATCTTCAAAATCTT 761
Db 7248 GGTAAATCTTCAAAATCTTCAAAATCTTCAAAATCTTCAAAATCTTCAAAATCTT 761
QY 761 GGTAAATCTTCAAAATCTTCAAAATCTTCAAAATCTTCAAAATCTTCAAAATCTT 821
Db 7308 GGTAAATCTTCAAAATCTTCAAAATCTTCAAAATCTTCAAAATCTTCAAAATCTT 821
QY 821 GGTAAATCTTCAAAATCTTCAAAATCTTCAAAATCTTCAAAATCTTCAAAATCTT 881
Db 7368 GGTAAATCTTCAAAATCTTCAAAATCTTCAAAATCTTCAAAATCTTCAAAATCTT 881

```

## RESULT 3

```

US-09-764-868-260
; Sequence 260, Application US/09764868
; Patent No. US2002016871A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT32
; CURRENT APPLICATION NUMBER: US/09/764,868
; PRIORITY FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 260
; LENGTH: 561
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (542)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (549)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-868-260

```

```

Query Match 10.2%; Score 97.2; DB 3; Length 561;
Best Local Similarity 52.2%; Pred. No. 1.1e-15;
Matches 235; Conservative 1; Mismatches 211; Indels 3; Gaps 1;
QY 282 GGTGTAACCAATCATGATCATCAATTTTCAAAATTTTCAAAATTTTCAAAATTTT 341

```

Db	104	GCTTACTACCTCAACCGTGATGCGGCCAGAGACCCCGACAGATTGCCAGCGACCCAGATCTC	153
Cy	312	ACGGGCGCTATTTCGACATTTGCGCGAAATTCGCAAAAATGCGCTCTGAAGCGTTCAAGGAGGA	401
Db	154	TGGCGTTACTGCGAATTTTGAAGCGCGTGCACCGCAAACTCGAGCGCGAATTCCGGRGCC	223
Cy	402	GATCGAAGACGTGGAGTTTCCCGAGGAAGACACTCAGTGGGACCTTCCTGAGAGATGAT	461
Db	224	AATGGCTGGCAATCTTCTTCCCGCGTAAAGCGCGTACCGCGAATTTTACTCGACAGACCAT	283
Cy	462	CTGTGAGCGTGGCGCGCGCGCTGCAGGAGTACCTGGGCTGCTTAGCGCATCCCGCTCCGT	521
Db	284	TGCGCGCGGTAGCGCGCGCGCTTTTGACGAGTTTTTGGGTACCTGCAAGCGAGTGGCTAGCT	343
Cy	522	GCGCGCGCTGCGCGAGTCTGCGAGCTTCCTACACCGGCGCGAGCTGGCGCAGGCTTTCGG	581
Db	344	GCGCGATCCCGCGGACCTCGAGAGACTTCTGTCTCCCGAGCTGGCGGCGGCACAGAG	403
Cy	582	CTGCTTCTGGCGCGATAGTACCTGGCGCGCTCGGAGTGTCTGTGGCGTGTGTCGCGT	641
Db	404	CCTCAGCTGTATGCGCTCTCTATCGTGAAGCGCTCTGGCAGTCTGGGCCAATGCTCGCGAGCT	463
Cy	642	GCGAGAGAAAGT----CAAGCGCACTGCGCTGCGCGCGCGCTCCCGGCGCTGTGCGCGCT	698
Db	464	GCAAGCTGAGTGTGGTACCGCGCTCTGGTGGAGAGCGCGCGCTGTGACCTGGCTGGGCT	523
Cy	699	GCTGCTTGGCGACCGCGAGCGTGGACCGCGCC	728
Db	524	GCGCGTGTGGTACCGGAGAACTGGAAAAAC	553

```

RESULT 4
US-09-853-753-1
: Sequence 1, Application US/09853753
: Publication No.: US200102669A1
: GENERAL INFORMATION:
: APPLICANT: Bech-Hansen, Torben
: TITLE OF INVENTION: GPI-Anchored Smal1 [Leucine-rich proteoglycan NYX
: FILE REFERENCE: 45499-2
: CURRENT APPLICATION NUMBER: US/00/853,753
: CURRENT FILING DATE: 2001-05-17
: PRIOR APPLICATION NUMBER: CA 2,106,241
: PRIOR FILING DATE: 2000-05-12
: NUMBER OF SEQ ID NOS: 14
: SOFTWARE: Patent in version 3.1
: SEQ ID NO 1
: LENGTH: 2297
: TYPE: DNA
: ORGANISM: Mus sp.
: PUBLICATION INFORMATION:
: AUTHORS: Bech-Hansen NT et al.
: TITLE: Mutations in NYX, encoding the leucine-rich proteoglycan nystroplon
: TITLE: Cause X-linked complete congenital stationary night blindness
: JOURNAL: Nature Genetics
: VOLUME: 26
: ISSUE: 3
: PAGES: 319-323
: DATE: 2000-11-01
: DATABASE ACCESSION NUMBER: GenBank / AF254868
: DATABASE ENTRY DATE: 2000-12-23
US-09-853-753-1

```

Query Match	5.7%;	Score 54.6;	DB 9;	Length 2297;
Best Local Similarity	51.2%;	Pred. No. 8.6e-05,		
Matches 154;	Conservative	0;	Mismatches 144;	Indels 3, Gaps 1

[illegible][illegible]

```

RESULT 5
US-09-815-242-4009
; Sequence 4009, Application US/03815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.01A
; CURRENT APPLICATION NUMBER: US/03/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 1410
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4009
; LENGTH: 7419
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-4009

```

Query Match	5.6%;	Score 53.4;	DB 10;	Length 7419;
Best Local Similarity	50.6%;	Pred. No. 0.0002;		
Matches 129, Conservative	0,	Mismatches 126,	Indels 0,	Gaps 0;

[illegible]

RESULT 7  
MS 6-991-495-115/c

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.

APPLICANT: Campos-Neto, Antonio

APPLICANT: Webb, John R.

APPLICANT: Dillon, Davin C.

APPLICANT: Skeiky, Yasir A.W.

APPLICANT: Bhatta, Ajay



RESULT 10  
US 08 957-425-28  
Sequence 28, Application US/08957425  
Publication No US20030069401A1  
GENERAL INFORMATION.  
APPLICANT: OPPELMANN, HERMANN  
OZKANAK, ENGIN  
KUBERASAMPATH, THIANGAVEL  
RUEGER, DAVID C.  
PANG, POY H. L.  
TITLE OF INVENTION: OSTEOGENIC DEVICES  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: TESTA, HIRWITZ & THIBEAULT  
STREET: 53 STATE STREET  
CITY: BOSTON  
STATE: MASSACHUSETTS  
COUNTRY: U.S.A.

```

1 21F: 02109
2 COMPUTER READABLE FORM:
3 MEDIUM TYPE: floppy disk
4 COMPUTER: IBM PC compatible
5 OPERATING SYSTEM: PC-DOS/MS-DOS
6 SOFTWARE: Patent in P-1-state #1 C, Version #1 25
7
8 CURRENT APPLICATION DATA:
9 APPLICATION NUMBER US/08/957,425
10 FILING DATE: 24-Oct-1997
11 CLASSIFICATION: <unknown>
12
13 PRIOR APPLICATION DATA:
14 APPLICATION NUMBER: US/08/447,570
15 FILING DATE: 21 FEB 1992
16 APPLICATION NUMBER: US 810,560
17 FILING DATE: 20-DEC-1991
18 APPLICATION NUMBER: US 827,052
19 FILING DATE: 28-JAN-1992
20 APPLICATION NUMBER: US 660,162
21 FILING DATE: 22-FEB-1991
22 APPLICATION NUMBER: US 621,988
23 FILING DATE: 04-DEC-1990
24 APPLICATION NUMBER: US 621,849
25 FILING DATE: 04-DEC-1990
26 APPLICATION NUMBER: US 616,374
27 FILING DATE: 21-NOV-1990
28 APPLICATION NUMBER: US 600,024
29 FILING DATE: 18-OCT-1990
30 APPLICATION NUMBER: US 599,543
31 FILING DATE: 18-OCT-1990
32 APPLICATION NUMBER: US 579,865
33 FILING DATE: 07-SEP-1990
34 APPLICATION NUMBER: US 569,920
35 FILING DATE: 20-AUG-1990
36 APPLICATION NUMBER: US 483,913
37 FILING DATE: 22-FEB-1990
38 APPLICATION NUMBER: US 422,613
39 FILING DATE: 17-OCT-1989
40 APPLICATION NUMBER: US 315,342
41 FILING DATE: 23-FEB-1989
42 APPLICATION NUMBER: US 232,630
43 FILING DATE: 15-AUG 1988
44 APPLICATION NUMBER: US 179,460
45 FILING DATE: 08-APR-1989
46
47 ATTORNEY/AGENT INFORMATION:
48 NAME: FITCHER, EDWARD R.
49 REGISTRATION NUMBER: 27,829
50 PREFERENCE/DOCKET NUMBER: CFF-201CP6
51 TELECOMMUNICATION INFORMATION:
52 TELEPHONE: 617/248-7000
53 TELEFAX: 617/248-7100
54 INFORMATION FOR SEQ ID NO: 28:
55 SEQUENCE CHARACTERISTICS:
56 LENGTH: 1723 base pairs
57 TYPE: nucleic acid
58 STRANDEDNESS: single
59 TOPOLOGY: linear
60 MOLECULE TYPE: cDNA
61 ORIGINAL SOURCE:
62 ORGANISM: Homo sapiens
63 TISSUE TYPE: HIPPOCAMPUS
64 FEATURE:
65 NAME/KEY: CDS
66 LOCATION: 490..1696
67 OTHER INFORMATION: /function= "OSTEOGENIC PROTEIN"
68 /note="hpf2 (cdna)"
69
70 SEQUENCE DESCRIPTION: SEQ ID NO: 28:
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

```

[illegible]

```

RESULT 11
US-09-939-964-1
: Sequence 1, Application US/09939964
: Publication No. US2003005452A1
: GENERAL INFORMATION:
: APPLICANT: Rosenthal, Andre
: APPLICANT: Freiberg, Christoph
: APPLICANT: Perret, Xavier Philippe
: APPLICANT: Broughton, William John
: TITLE OF INVENTION: Genomic Sequence of Phicobium SP. NCR 234 Symbiotic
: FILE REFERENCE: CAP0008
: CURRENT APPLICATION NUMBER: US/09/939,964
: CURRENT FILING DATE: 2001-09-27
: PRIOR APPLICATION NUMBER: 09/214,808
: PRIOR FILING DATE: 1999-06-22
: NUMBER OF SEQ ID NOS: 1
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 1
: LENGTH: 536155
: TYPE: DNA
: ORGANISM: Phicobium
: US-09-939-964-1

```

[illegible]

QY 808 CTGGACGCCATGTCCTCCGCCTAGCGCCTGGGCAAGACTTGT 854  
| | | | |  
Db 20622 GCGCACCCATTGGCACTGTTTGCCGGCGGTGTGGCAGCAGCGGCTTCGT 20668

```

RESULT 12
US-09-860-846-1
: Sequence 1, Application US/09860846
: Patent No. US2002164742A1
: GENERAL INFORMATION:
: APPLICANT: Sherman, D.H.
: APPLICANT: Liu, H.
: APPLICANT: Xue, Y.
: APPLICANT: Zhao, L.
: TITLE OF INVENTION: DNA encoding methymycin and pikromycin
: FILE REFERENCE: 600.430US1
: CURRENT APPLICATION NUMBER: US/2/560.846
: CURRENT FILING DATE: 2001-05-18
: PRIOR APPLICATION NUMBER: 09/105,537
: PRIOR FILING DATE: 1998-06-26
: NUMBER OF SEQ ID NOS: 43
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 1
: LENGTH: 15872
: TYPE: DNA
: ORGANISM: Streptomyces venezuelae
US-09-860-846-1

```

[illegible]

```

1 RESULT 13
2 US-09-861-289-1
3 : Sequence 1, Applicant: us/c9861-289
4 : Patent No. US20020110897A1
5 : GENERAL INFORMATION:
6 : APPLICANT: Sherman, D.H.
7 : APPLICANT: Liu, H.
8 : APPLICANT: Xue, Y.
9 : APPLICANT: Zhao, L.
10 : TITLE OF INVENTION: DNA encoding methymycin and pikromycin
11 : FILE REFERENCE: 600.438US1
12 : CURRENT APPLICATION NUMBER: US/09/861,289
13 : CURRENT FILING DATE: 2001-05-18
14 : PRIOR APPLICATION NUMBER: 09/105,537
15 : PRIOR FILING DATE: 1998-06-26
16 : NUMBER OF SEQ. ID NOS: 43
17 : SOFTWARE: FaastSD for Windows Version 3.0
18 : SEQ. ID NO. 1
19 : LENGTH: 15872
20 : TYPE: DNA

```

ORGANISM: Streptomyces venezuelae  
US-09-861-289-1

Query Match 5.3%; Score 50.2; DB 10, Length 15872;  
Best Local Similarity 49.1%; Pred. No. 0.0014;  
Matches 113, Conservative 0, Mismatches 138, Indels 0, Gaps 0;

QY 546 CTTCTTACCGCGCGGAGCTTGGCGAGCTTTGGCTCTCTCGCGGCGGCGAGTACCC 605  
DB 9016 CTTCTTACCGCGCGGAGCTTGGCGAGCTTTGGCTCTCTCGCGGCGGCGAGTACCC 605  
QY 606 GCGCGCTTCTTACCGCGCGGAGCTTGGCGAGCTTTGGCTCTCTCGCGGCGGCGAGTACCC 665  
DB 9076 GCGCGCTTCTTACCGCGCGGAGCTTGGCGAGCTTTGGCTCTCTCGCGGCGGCGAGTACCC 665  
QY 666 CTTCTTACCGCGCGGAGCTTGGCGAGCTTTGGCTCTCTCGCGGCGGCGAGTACCC 725  
DB 9136 CTTCTTACCGCGCGGAGCTTGGCGAGCTTTGGCTCTCTCGCGGCGGCGAGTACCC 725  
QY 726 CTTCTTACCGCGCGGAGCTTGGCGAGCTTTGGCTCTCTCGCGGCGGCGAGTACCC 785  
DB 9196 CTTCTTACCGCGCGGAGCTTGGCGAGCTTTGGCTCTCTCGCGGCGGCGAGTACCC 785  
QY 786 CTTCTTACCGCGCGGAGCTTGGCGAGCTTTGGCTCTCTCGCGGCGGCGAGTACCC 845  
DB 9256 CTTCTTACCGCGCGGAGCTTGGCGAGCTTTGGCTCTCTCGCGGCGGCGAGTACCC 845

## RESULT 14

US 10-076-157-7  
Sequence 7, Application US/10076157  
Publication No. US20010027109A1  
GENERAL INFORMATION:

APPLICANT: Pompeius, Markus  
APPLICANT: Suelberger, Harald  
APPLICANT: Joellken, Hans Wolfgang  
APPLICANT: Dostal, Jose Luis Revuelta  
APPLICANT: Jimenez, Alberto  
TITLE OF INVENTION: Genes of purine biosynthesis from Ashbya gossypii and the use thereof  
TITLE OF INVENTION: in  
FILE REFERENCE: 48684DIV  
CURRENT APPLICATION NUMBER: US/10/076.157  
CURRENT FILING DATE: 2002-02-15  
PRIOR APPLICATION NUMBER: US 09/212,247  
PRIOR FILING DATE: 1998-12-16  
NUMBER OF SEQ ID NOS: 21  
SOFTWARE: WordPerfect V. 6.1  
SEQ ID NO 7  
LENGTH: 3616  
TYPE: DNA  
ORGANISM: Ashbya gossypii  
FEATURE:  
NAME/KEY: CDC  
LOCATION: 864..1316  
NAME/KEY: CDC  
LOCATION: 1478..2592  
US 10-076-157-7

Query Match 5.3%; Score 50, DB 9, Length 3616;  
Best Local Similarity 45.9%; Pred. No. 0.0013;  
Matches 170, Conservative 0, Mismatches 200, Indels 0, Gaps 0;

QY 474 GCGCGCTTCTTACCGCGCGGAGCTTGGCGAGCTTTGGCTCTCTCGCGGCGGCGAGTACCC 533  
DB 224 GCGCGCTTCTTACCGCGCGGAGCTTGGCGAGCTTTGGCTCTCTCGCGGCGGCGAGTACCC 533  
QY 534 GCGCGCTTCTTACCGCGCGGAGCTTGGCGAGCTTTGGCTCTCTCGCGGCGGCGAGTACCC 593  
DB 284 GCGCGCTTCTTACCGCGCGGAGCTTGGCGAGCTTTGGCTCTCTCGCGGCGGCGAGTACCC 593  
QY 594 GCGCGCTTCTTACCGCGCGGAGCTTGGCGAGCTTTGGCTCTCTCGCGGCGGCGAGTACCC 653  
DB 653 GCGCGCTTCTTACCGCGCGGAGCTTGGCGAGCTTTGGCTCTCTCGCGGCGGCGAGTACCC 653

DB 344 CTTCTTACCGCGCGGAGCTTGGCGAGCTTTGGCTCTCTCGCGGCGGCGAGTACCC 403  
QY 654 CTTCTTACCGCGCGGAGCTTGGCGAGCTTTGGCTCTCTCGCGGCGGCGAGTACCC 713  
DB 404 CTTCTTACCGCGCGGAGCTTGGCGAGCTTTGGCTCTCTCGCGGCGGCGAGTACCC 463  
QY 714 CTTCTTACCGCGCGGAGCTTGGCGAGCTTTGGCTCTCTCGCGGCGGCGAGTACCC 773  
DB 464 CTTCTTACCGCGCGGAGCTTGGCGAGCTTTGGCTCTCTCGCGGCGGCGAGTACCC 773  
QY 774 CTTCTTACCGCGCGGAGCTTGGCGAGCTTTGGCTCTCTCGCGGCGGCGAGTACCC 833  
DB 524 CTTCTTACCGCGCGGAGCTTGGCGAGCTTTGGCTCTCTCGCGGCGGCGAGTACCC 833  
QY 834 CTTCTTACCGCGCGGAGCTTGGCGAGCTTTGGCTCTCTCGCGGCGGCGAGTACCC 893  
DB 584 CTTCTTACCGCGCGGAGCTTGGCGAGCTTTGGCTCTCTCGCGGCGGCGAGTACCC 893

## RESULT 15

US 09-796-679-5/c  
Sequence 5, Application US/09796679  
Publication No. US20010013076A1  
GENERAL INFORMATION:

APPLICANT: Robinson, Anthony J  
APPLICANT: Lytle, David J  
TITLE OF INVENTION: Parapoxvirus vectors  
FILE REFERENCE: 23607 MPB  
CURRENT APPLICATION NUMBER: US/09/796,679  
CURRENT FILING DATE: 2001-03-30  
PRIOR APPLICATION NUMBER: 09/155,421  
PRIOR FILING DATE: 1998-09-29  
PRIOR APPLICATION NUMBER: PCT/NZ97/00040  
PRIOR FILING DATE: 1997-03-27  
PRIOR APPLICATION NUMBER: NZ 286284  
PRIOR FILING DATE: 1996-03-29  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: Patentr Ver. 2.1  
SEQ ID NO 5  
LENGTH: 4020  
TYPE: DNA  
ORGANISM: Oxf virus strain NZ-2  
US-09-796-679-5

Query Match 5.2%; Score 49.6; DB 9, Length 4020;  
Best Local Similarity 50.8%; Pred. No 0.0017;  
Matches 151, Conservative 0, Mismatches 149, Indels 2, Gaps 1;

QY 514 CTTCTTACCGCGCGGAGCTTGGCGAGCTTTGGCTCTCTCGCGGCGGCGAGTACCC 573  
DB 1070 CTTCTTACCGCGCGGAGCTTGGCGAGCTTTGGCTCTCTCGCGGCGGCGAGTACCC 573  
QY 574 CTTCTTACCGCGCGGAGCTTGGCGAGCTTTGGCTCTCTCGCGGCGGCGAGTACCC 633  
DB 1010 CTTCTTACCGCGCGGAGCTTGGCGAGCTTTGGCTCTCTCGCGGCGGCGAGTACCC 633  
QY 634 CTTCTTACCGCGCGGAGCTTGGCGAGCTTTGGCTCTCTCGCGGCGGCGAGTACCC 693  
DB 693 CTTCTTACCGCGCGGAGCTTGGCGAGCTTTGGCTCTCTCGCGGCGGCGAGTACCC 693  
QY 694 CTTCTTACCGCGCGGAGCTTGGCGAGCTTTGGCTCTCTCGCGGCGGCGAGTACCC 753  
DB 694 CTTCTTACCGCGCGGAGCTTGGCGAGCTTTGGCTCTCTCGCGGCGGCGAGTACCC 753  
QY 754 CTTCTTACCGCGCGGAGCTTGGCGAGCTTTGGCTCTCTCGCGGCGGCGAGTACCC 813  
DB 813 CTTCTTACCGCGCGGAGCTTGGCGAGCTTTGGCTCTCTCGCGGCGGCGAGTACCC 813  
QY 814 CTTCTTACCGCGCGGAGCTTGGCGAGCTTTGGCTCTCTCGCGGCGGCGAGTACCC 873  
DB 772 CTTCTTACCGCGCGGAGCTTGGCGAGCTTTGGCTCTCTCGCGGCGGCGAGTACCC 873



Tue Apr 22 05:41:53 2003

us-09-816-697a-1.rnpb

Page 9

Search completed: April 20, 2003, 20:19:49  
Job time : 664 secs

---





RESULT 2  
US 08-804-227C-7  
Sequence 7, Application US/08804227C  
Patent No. 5876991

```

1 LOCATION: 350..14062
2 FEATURE:
3 NAME/KEY: CDS
4 LOCATION: 14046..20036
5 FEATURE:
6 NAME/KEY: CDS
7 LOCATION: 20110..31284
8 FEATURE:
9 NAME/KEY: CDS
10 LOCATION: 31329..36071
11 FEATURE:
12 NAME/KEY: CDS
13 LOCATION: 36155..41830
14 US-08-804-227C-7
15
16 Query Match          5.9%; Score 56.2; DB 2; Length 44377;
17 Best Local Similarity 47.8%; Pval: N= 0.0044;
18 Matches 163; Conservative 0; Mismatches 178; Indels 0; Gaps 0;
19
20 QY      508 GCCATCGGCTGGTTCAGTGACGTGGTTCCGCGGCAAGTTCTTGAACATTCTTCAAGCGCGCGGCAATG 567
21         ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
22 Db      30925 GTCACTGCCGCGGCTGAGGTTCCTCCGCGGCAAGTTCTTGAACATTCTTCAAGCGCGCGGCAATG 4984
23
24 QY      568 CGCAGAGCTTTTCGCGTGCCTGCTGCGGCGCGGCAAGTAACGCGCGCGCGCGCGCGCGCGCGCGT 627
25         ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
26 Db      30945 TCTCTCGCATCTTGAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGT 31044
27
28 QY      628 CGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGT 687
29         ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
30 Db      31045 TCGCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGT 31104
31
32 QY      688 CTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGT 747
33         ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
34 Db      31105 CTGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGT 31164
35
36 QY      748 GGAAGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGT 807
37         ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
38 Db      31165 GGTCTTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGT 31224
39
40 QY      808 TTGAACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGT 848
41         ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
42 Db      31225 AGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGT 31265
43
44 RESULT 3
45 US-08-804-198-1
46 Sequence 1, Application US/08804198
47 Patent No. 5945320
48 GENERAL INFORMATION:
49 APPLICANT: Burgett, Stanley G.
50 APPLICANT: Kuhstoss, Stuart A.
51 APPLICANT: Rao, Nagaraja P
52 APPLICANT: Richardson, Mark A.
53 APPLICANT: Postek, Paul R., Jr.
54 TITLE OF INVENTION: PLATENOIDE SYNTHASE GENE
55 NUMBER OF SEQUENCES: 6
56 CORRESPONDENCE ADDRESS:
57 ADDRESSEE: PAUL R. CANTRILL 1138
58 STREET: LILLY CORPORATE CENTER
59 CITY: INDIANAPOLIS
60 STATE: IN
61 COUNTRY: USA
62 ZIP: 46285
63 COMPUTER READABLE FORM:
64 MEDIUM TYPE: Floppy disk
65 COMPUTER: Macintosh
66 OPERATING SYSTEM: Macintosh 7.0
67 SOFTWARE: Microsoft Word 5.1
68 CURRENT APPLICATION DATA:
69 APPLICATION NUMBER: US/08/804,198
70 FILING DATE:
71 CLASSIFICATION: 435
72 ATTORNEY/AGENT INFORMATION:

```













```

; TELEFAX: (508) 435-6951
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 884 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..884
; OTHER INFORMATION: /note= "hnp-2 genomic sequence"
; FEATURE:
; NAME/KEY: exon
; LOCATION: 1..837
; OTHER INFORMATION: /note= "EXON ONE"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 884
; OTHER INFORMATION: /note= "A GAP OCCURS BETWEEN
; OTHER INFORMATION: POSITIONS 884 IN THIS SEQUENCE AND POSITION 1 IN SEQ ID NO 14"
; US 09 219-391-11

```

```

Query Match          5.5%; Score 52.2; DB 3; Length 884;
Host Local Similarity 48.8%; Pred. No. 0.016;
Matches 141; Conservative 0; Mismatches 148; Indels 0; Gaps 0;

QY 472 TTTTGGGCTTGTATGATATGCTTGGGCTTGTATGATATGCTTGGGCTTGGCTCC 531
Db 495 TTTTGGGCTTGTATGATATGCTTGGGCTTGTATGATATGCTTGGGCTTGGCTCC 554

QY 532 GATGATTTTGTGATTTTGTATGATATGCTTGGGCTTGGCTTGGCTTGGCTGG 591
Db 555 CTGGGCTTGTGATTTTGTATGATATGCTTGGGCTTGGCTTGGCTTGGCTGG 614

QY 592 GCCGGCTAGTACCGCGCGCTTGTATGATATGCTTGGGCTTGGCTTGGCTGG 651
Db 615 GCTGGCTAGTACCGCGCGCTTGTATGATATGCTTGGGCTTGGCTTGGCTGG 674

QY 652 CTATGCGCTATGCTTGTATGATATGCTTGGGCTTGGCTTGGCTTGGCTGG 711
Db 675 CTTGGCTATGCTTGTATGATATGCTTGGGCTTGGCTTGGCTTGGCTGG 734

QY 712 CTATGCTATGCTTGTATGATATGCTTGGGCTTGGCTTGGCTTGGCTGG 760
Db 735 CTGATGCTTGTATGATATGCTTGGGCTTGGCTTGGCTTGGCTGG 783

```

Search completed: April 20, 2003, 17:03:16  
Job time: 290 secs

GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: April 20, 2003, 14:02:01, Search time 45 Seconds  
(without alignments) 675.077 Million cell updates/sec

Title: US-09-816-697A-2

Sequence: 1 MASPEHSGSGCMGPICTCT... PPTPTPTITKELTVREYLH 316

Scoring table: RUCSUM62

Gapop 10.0, Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: PIR 73: \*  
1: Pirt.\*  
2: Pirt.\*  
3: Pirt.\*  
4: Pirt.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	117	7.1	952	2 E84534	hypothetical prote
2	116	7.0	1010	2 T41077	hypothetical struc
3	111	6.7	473	2 I50420	nonerythrocid alpha
4	111	6.7	1104	2 T49735	related to MDN1 pr
5	111	6.7	2477	1 STCHA	spectrin alpha cha
6	102.5	6.2	1127	2 S47445	MDM1 protein - yea
7	102	6.2	1030	2 A32612	spectrin alpha cha
8	101	6.1	1325	2 T14760	hypothetical prote
9	99.5	6.0	705	2 A35363	synapsin I splice
10	99	6.0	475	2 A27671	spectrin alpha cha
11	99	6.0	2472	2 A35715	foxlin alpha chain
12	98	5.9	459	2 T31608	hypothetical prote
13	97.5	5.9	420	2 T08621	hypothetical prote
14	97	5.9	4684	2 A59404	plectin (imported)
15	96	5.8	441	2 C03318	conserved hypochet
16	96	5.8	4574	2 C02520	plectin - human
17	95.5	5.8	373	2 C02546	two-component syst
18	95.5	5.8	398	2 S75202	hypothetical prote
19	95	5.8	540	2 T40939	probable Ca-aimed
20	95	5.8	943	2 A82934	exonuclease ABC c
21	94.5	5.7	1092	2 H81982	hypothetical prote
22	93	5.7	578	2 T31345	hypothetical phox
23	93	5.6	401	2 T27671	spectrin alpha cha
24	93	5.6	475	2 T35985	probable large pro
25	92.5	5.6	814	2 T05537	probable serine/th
26	92	5.6	295	2 T42515	hypothetical prote
27	92	5.6	706	2 E30411	synapsin Ia - bovi
28	92	5.6	1033	2 T42701	hypothetical prote

30	91.5	5.5	433	2 S19996	hypothetical prote
31	91	5.5	448	2 D75197	tRNA nucleotidyltr
32	91	5.5	640	2 S67656	hypothetical prote
33	91	5.5	892	2 T29420	probable transcrip
34	90.5	5.5	143	2 T40731	probable golgi mem
35	90.5	5.5	297	2 H72670	hypothetical prote
36	90.5	5.5	472	2 D70708	probable purp prot
37	90.5	5.5	533	2 T75583	GSDER family prote
38	90	5.5	545	2 B70374	conserved hypochet
39	90	5.5	1446	1 A45344	immediate-early pr
40	90	5.5	1460	1 EDBE1F	retinoic acid rece
41	89.5	5.4	455	2 S13512	conserved hypochet
42	89	5.4	1162	2 D83454	plectin - rat
43	89	5.4	4687	1 A39638	hypothetical prote
44	88.5	5.4	950	2 F86286	plectin
45	88.5	5.4	2591	2 T30288	pistinnacytin I sy

## ALIGNMENTS

### RESULT 1

E84534

hypothetical protein At2g15900 (imported) - Arabidopsis thaliana

C>Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02 Feb 2001

C/Accession: E84534

P/lin: X; Kaul, S.; Pounsley, S. D.; Shea, T. P.; Benito, M. J.; Town, C. D.; Fujii, C. Y.;

M.; Koo, H.; Moffat, K. S.; Cronin, L. A.; Shen, M.; Vanaken, S. E.; Unayam, L.; Tallon, L.;

Bus, D.; Nierman, W. C.; White, O.; Eisen, J. A.; Salzberg, S. L.; Fraser, C. M.; Venter, J.

Nature 402, 761-768, 1999

A>Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A/Reference number: A84420; NCID:20083487; FWD:10617197

A/Accession: E84534

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-952 <STO>

A/Cross-references: GB: A800203; NID: g5106243; RID: A041976.1; GSFB: G000139

A/Genes: A:Gene: At2g15900

A/Map position: 2

Query Match 7.1%; Score 117; DB 2; Length 952;

Best local similarity 26.3%; Pred. No. 0.25; Mismatches 70; Indels 20; Gaps 5;

Matches 41; Conservative 25; Mismatches 70; Indels 20; Gaps 5;

QY 42 HLDTHGSLSSMTRELDQYQWQKCPWFHKLFEIASPIEEPVSFVYUUIVI 101

Db 503 HNDNYISQVSGSVLHKEGCHLFCR ..VLGAVFEQSGKFAVSIATV 553

QY 102 QTSFENKAVLEFYSQAKQKALIKTFEEIE--EVEFPPIHTQNFARFMGEP 159

Db 554 DV--ENKTFVFRPVSNEFPLHPLQ----KEIPNYQLPEPKFISSSTELAFVHRPC 605

QY 159 RALQFYLLLYALPVCVPSPFELDFTRPELRBAG 194

Db 606 IQLDKYLDLLCLIANVABQHEVWUFLSASKVYSFG 641

RESULT 2

T41077

hypothetical structural protein - fission yeast (Schizosaccharomyces pombe)

C/Species: Schizosaccharomyces pombe

C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999

C/Accession: T41077

P/McDonnell, P. C.; Rajandream, M. A.; Barrell, B. G.; Zimmermann, W.; Wambutt, R.

submitted to the EMBL Data Library, August 1999

A/Reference number: 221822

A/Accession: T41077

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-1010 <MCD>

A/Cross references: EMBL: A109357; FICD: CAB530/6.1; GSFB: G000068; SFDB: SP0016A11.04



F:11338-1443/Domain: spectrin/dystrophin repeat homology <SP12>  
 F:11444-1549/Domain: spectrin/dystrophin repeat homology <SP13>  
 F:11500-1661/Domain: spectrin/dystrophin repeat homology <SP14>  
 F:11662-1767/Domain: spectrin/dystrophin repeat homology <SP15>  
 F:11768-1873/Domain: spectrin/dystrophin repeat homology <SP16>  
 F:11874-1979/Domain: spectrin/dystrophin repeat homology <SP17>  
 F:11980-2086/Domain: spectrin/dystrophin repeat homology <SP18>  
 F:12095-2200/Domain: spectrin/dystrophin repeat homology <SP19>  
 F:12209-2315/Domain: spectrin/dystrophin repeat homology <SP20>  
 F:12328-2360/Domain: spectrin/dystrophin repeat homology <SP21>  
 F:12371-2403/Domain: calmodulin repeat homology <EF2>

Query Match 6.7% Score 111; DB 1; Length 2477;  
 Best Local Similarity 21.2%; Pred. No. 2.4;  
 Matches 66; Conservative 44; Mismatches 134; Indels 68; Gaps 8;

QY 37 PGPDGHLTHSGLSNSSMTTPELQYVWQNCRCRWKHYVLLFETASAPTEEP-KVSFV-94  
 DB 1823 PAIQGLDGTGKSLSDNTIGKEIQRLAQFVDHMKELKQLAARGQLLESLEYQGFVA 1882  
 QY 95 -----VQIIVIQGTFDNKNAVLPRYSDFAKLQVAL 128  
 DB 1843 NVEEEMAMINEKMTLVASEDYQDTLAAIGLKHAEFTFTVHKDFVNDVCANGEDLI 1942  
 QY 129 KTFPEIEDVEFPFPHLTGNFA--EEMTCFPPPALQEVLLGLVAIPVPSPEFLDPLTR 186  
 DB 1943 KKNHHVENITAKMKGLGKVSLEKAAAKRKAIDE-----NSAFIQFPMK 1989  
 QY 187 PELRPAF-----GCLPAGQPPALVELLPLPLEKLTANCPA-----AAVPALCAVLL 235  
 DB 1990 ADVESWIKGEKENSILKTDYDPDLSVQTLTKQETFCAGLQAFQEGEINATLTKPGL 2049  
 QY 236 C-----HFDLPPPAFAFAAGPALCP--LQAPGSHRYVAPLLDAMVPLAVLG 291  
 DB 2060 AAKHIQSAITEVPASLKMKNQLLANSAPAKKLEAQEHFRKVEDLFTFAKASAFN 2109  
 QY 282 KDFVTLQERLEE 293  
 DB 2110 SWFENAEDITD 2121

RESULT 6  
 S47445  
 MDML protein - yeast (Saccharomyces cerevisiae)  
 N:Alternate names: protein 0.13; protein YM8339 15c; protein YML104c  
 C:Species: Saccharomyces cerevisiae  
 C>Date: 02-Dec-1994 #sequence\_revision 06-Sep-1996 #text\_change 21-Jul-2000  
 C:Accession: S47445; S53968; A42636; S51916  
 R:Barrell, B.G.  
 submitted to the EMBL Data Library, August 1994  
 A:Reference number: S47445  
 A:Accession: S47445  
 A:Molecule type: DNA  
 A:Residues: 1-676 <BAR>  
 A:Cross-references: EMBL:X80835; NID:G530339; PID:G530340; MIPS:YML104c  
 R:McConnell, S.J.; Yaffe, M.P.  
 J. Cell Biol. 118, 385-395, 1992  
 A>Title: Nuclear and mitochondrial inheritance in yeast depends on novel cytoplasmic str  
 A:Reference number: A42636; MUID:J533595; PMID:1378448  
 A:Accession: A42636  
 A:Molecule type: DNA  
 A:Residues: 688-771; 'P', 773-898; 'I', 900-1062; 'K', 1064-1080; 'S', 1082-1095; 'R', 1097-1127  
 A:Cross-references: EMBL:X66371; NID:G33915; PID:CAA47014.1; PID:G33916  
 A:Name: sequence extracted from NCBI database (NCBI:108433, NCBI:108434)  
 R:Strilling, C.J.  
 submitted to the EMBL Data Library, February 1995

A:Reference number: S51916  
 A:Accession: S51916  
 A:Molecule type: DNA  
 A:Residues: 874-1127 <ST1>  
 A:Cross-references: EMBL:X65783; NID:G671637; PID:G671638  
 A:Gene: SGD:MDM1  
 A:Cross-references: SGD:SGC04572; MIPS:YML104c  
 A:Map position: 13L

Query Match 6.2% Score 102.5; DB 2; Length 1127;  
 Best Local Similarity 26.3%; Pred. No. 4.4;  
 Matches 49; Conservative 30; Mismatches 52; Indels 55; Gaps 11;

QY 42 HLDTHSGLSN-----SSMTREL--QYV--QNCRCRWKHYVLLF-----78  
 DB 733 HLITVADLTNMGCLKILYSQFTLKELEWELLKQVWQVNGNSLFPRTYTYTFEGVF 740  
 QY 79 -ELASARIEPKSKFVYQIIVY-----QTSFDNNAVLPRYSDFAKLQKALLTF 131  
 DB 791 SENSNGKKE-----ITYIINHFNNGVSWD-----MAFPNEFFELNTLYKRYF 839  
 QY 132 FE--EIEDVEFPFK-----HGTGFAEMTCFPPALQEVLLGLVAIPVPSPEFL 181  
 DB 840 RDLMPQLDL-FPSKVMGLKHYVTLTYL--PPKQLEYLPELLSISITCEDNIFP 895  
 QY 182 DFLTRP 187  
 DB 896 RFLTRP 901

RESULT 7  
 A32612  
 spectrin alpha chain, nonerythroid - rat (fragment)  
 C:Species: Rattus norvegicus (Norway rat)  
 C>Date: 21-May-1990 #sequence\_revision 21 May-1990 #text\_change 29-Sep-1999  
 C:Accession: A32612  
 R:Hong, W.; Doyle, D.  
 J. Biol. Chem. 264, 12758-12764, 1989  
 A>Title: Cloning and analysis of cDNA clones for rat kidney alpha-spectrin.  
 A:Reference number: A32612; MUID:89327227; PMID:2753883  
 A:Accession: A32612  
 A:Molecule type: mRNA  
 A:Residues: 1-1030 <HON>  
 A:Cross-references: GB:J294828; NID:G279313; PID:AAA46770.1; PID:G279314  
 C:Superfamily: spectrin alpha chain, calmodulin repeat homology; SH homology; spectrin/  
 C:Keywords: actin binding; cytoskeleton; EF hand  
 F:47-152/Domain: spectrin/dystrophin repeat homology <SP12>  
 F:799-904/Domain: spectrin/dystrophin repeat homology <SP19>  
 F:913-1019/Domain: spectrin/dystrophin repeat homology <SP20>

Query Match 6.2% Score 102; DB 2; Length 1030;  
 Best Local Similarity 20.4%; Pred. No. 4.4;  
 Matches 64; Conservative 46; Mismatches 132; Indels 72; Gaps 8;

QY 37 PGPDGHLTHSGLSNSSMTTPELQYVWQNCRCRWKHYVLLFETASAPTEEP-KVSFV-94  
 DB 527 PAIQGLDGTGKSLSDNTIGKEIQRLAQFVDHMKELKQLAARGQLLESLEYQGFVA 586  
 QY 95 -----VQIIVIQGTFDNKNAVLPRYSDFAKLQVAL 128  
 DB 1843 NVEEEMAMINEKMTLVASEDYQDTLAAIGLKHAEFTFTVHKDFVNDVCANGEDLI 1942  
 QY 129 KTFPEIEDVEFPFPHLTGNFA--EEMTCFPPPALQEVLLGLVAIPVPSPEFLDPLTR 186  
 DB 1943 KKNHHVENITAKMKGLGKVSLEKAAAKRKAIDE-----NSAFIQFPMK 1989  
 QY 187 PELRPAF-----GCLPAGQPPALVELLPLPLEKLTANCPA-----AAVPALCAVLL 235  
 DB 1990 ADVESWIKGEKENSILKTDYDPDLSVQTLTKQETFCAGLQAFQEGEINATLTKPGL 2049  
 QY 236 C-----HFDLPPPAFAFAAGPALCP--LQAPGSHRYVAPLLDAMVPLAVLG 291

Dh 754 AANH10--SKALEAPHASLMRPWTGLANSATPRYKLLFAGSHPRPVNLELTPAYASA 811

Qy 280 LGKLPVLAEPLEE 293

Dh 812 FNSWFENAEEDLTD 825

# RESULT 8

T14790 hyperthermal protein NP2564P0562 : human (fragment)

C:Species: Homo sapiens (man)

C:Date: 20-Sep-1999 #sequence revision 20-Sep-1999 #text\_change 09 Jun 2000

C:Accession: T14790

R:Blum, H.; Baerendsch, S.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S. submitted to the Protein Sequence Database, August 1999

A:Reference number: 218179

A:Accession: T14790

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 11325 <BLU>

A:Cross-references: EMBL, AL110273

A:Experimental source: fetal brain; clone DKFZ564P0562

C:Genetics:

A:Note: DKFZ564P0562.1

C:Superfamily: spectrin alpha chain, calmodulin repeat homology, SH3 homology; spectrin/

C:Keywords: EF hand

## Query Match

Best Local Similarity 20.8%, Pred No. 71; Matches 65; Conservative 43; Mismatches 136; Indels 68; Gaps 8;

```

Qy 37 PGDGLDTHSGLSNSSMTTRELQYWMQNCQPMWVVLFEIASAPITFPF-KVSKFV- 94
Dh 807 PAIGVLTCTPKKISDNTIGKEITQPLAQPVHMKELQLAANAGQPLESLEYQFVA 866
Qy 95 -----VQIIVIGTGSFQNNKAVLERYSDFAKLQKAL 128
Dh 867 NVEEBAMINEKMTLVASEYGDTLAIOGLLKHEAFETDFTHKDVNDVCTHGOGLI 926
Qy 129 KTFPPFIEPVPPPHLTENFA--EEMICEPRALOEYIGLXLA-RCVRRSREFDELTR 186
Dh 927 KYNHHFENISSMKPLNHYVSNIFPDAAQPRAYLDE...--TSAPLQEFKWK 973
Qy 187 PELREAF-----GCLPAGQVPYALVELLPLVLEKLTANCPA-----AAVPALCAVL 235
Dh 974 ADVAFSWIGIKRNSLTDDYGRPLSSVQTLTKQEPFADGICAPQCEGIANITLAKOQL 1033
Qy 236 C-----HPIIDPPAEFAAGERALQF--LQAREGHRYVAPLLDAMVRLVALG 281
Dh 1034 AANHVSOKALEAPHACLMPPWSQILANSAPRYKTLFAGSHPRPVNLELTPAYASA 1093
Qy 282 KRFVTIQFPLEE 293
Dh 1094 SWFENAEEDLTD 1105

```

# RESULT 9

A35363 synapsin I splice form a - human

C:Species: Homo sapiens (man)

C:Date: 17-Aug-1990 #sequence revision 17-Aug-1990 #text\_change 01 Dec 2000

C:Accession: A35363

R:Stuehrhof, T.C.

J. Biol. Chem. 265, 7849-7852, 1990

A:Title: The structure of the human synapsin I gene and protein.

A:Reference number: A35363; MUID:90243651; PMID:2110562

A:Accession: A35363

A:Molecule type: DNA

A:Residues: 1-705 <SUE>

A:Cross-references: GB:M58371; GB:J05431

A:Accession: A35363

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-659, 'KASPAQACQ' <SUE>

A:Cross-references: GB:M58378; GB:J05431

R:Saurwald, A.; Hoeseche, C.; Oschwald, F.; Kilimann, M.W.

J. Biol. Chem. 265, 14932-14937, 1990

A:Title: The 5'-flanking region of the synapsin I gene. A G+C-rich, TATA- and CAAT-less,

A:Reference number: A35805; MUID-0036667; PMID:2118519

A:Accession: A35805

A:Molecule type: DNA

A:Residues: 1-125 <SAU>

A:Cross-references: GB:M58301; NID:933865; PIR:MAA6668.1; PIR:J03364; GB:J05630

C:Genetics:

A:Gene: GDB:SYN1

A:Cross-references: GDB:119606; OMIM:313440

A:Map position: Xp11.23-Xp11.23

C:Keywords: actin binding; alternative splicing; phosphoprotein

## Query Match

Best Local Similarity 25.0%; Pred No. 44; Matches 72; Conservative 29; Mismatches 96; Indels 91; Gaps 17;

```

Qy 4 PEHGC--SPGCM-GPICTARTQCEAPATGULPHPGIPGHLTHSGLSNSSMTTREL 60
Dh 32 PPPGASHPPGATPGGTATPARRSGVAPAAAPSPSGSGG3GFSSLSNAVKQT--- 88
Qy 61 QCYWMQNCQPMWVVLFEIASAPILEPVSKFVVVQIIVIGTGSFQNNKAVLERYSDF 112
Dh 89 -----MAAATFSEQ-----VGGSGGAGCGAASRVLLV 118
Qy 113 LERRYDFPAFLKFLKTPPEI---EVEFIP- PHLTNFAEEMICEPRALOEY 164
Dh 119 IDEPHTDAKPKFG--KIHGEIDIKVQAEFSDLNVAANAGQPLESLEYQFVA 170
Qy 165 LSLYAIKVPKSELELF-LTRGHLKFAVQCIIPAKYVPALVELLPLVLEKLTANCP 223
Dh 171 -----GVKVPKSLKP--DFVLIK--QHAFSMARKVDY-PSL-----VIGLQ----- 206
Qy 224 AAAPALCAVLTPHPLDPPAEFAAGERALQFQAECHRYVAPLLD 271
Dh 207 YAGISVNSLSHVNFCDKP WFAQWVPLHKLTGE-----FPLID 248

```

# RESULT 10

A27671 spectrin alpha chain, nonerythroid - rat (fragment)

C:Species: Rattus norvegicus (Norway rat)

C:Date: 19 May 1987 #sequence revision 13 May 1988 #text\_change 29-Sep-1999

C:Accession: A27671

R:Leto, T.L.; Fortugno-Erikson, D.; Barton, D.; Yang-Peng, T.L.; Francke, U.; Harris, A.S.

Mol. Cell. Biol. 8, 1-9, 1988

A:Title: Comparison of nonerythroid alpha spectrin genes reveals strict homology among d

A:Reference number: A9397; MUID-88094373; PMID-3336352

A:Accession: A27671

A:Molecule type: mRNA

A:Residues: 1-475 <LET>

A:Cross-references: GB:M19726; NID:920541; PIR:MAA4177.1; PIR:9205642

A:Note: The authors translated the codon GGC for residue 18 as Ser, GAG for residue 45 as

Arg, AGA for residue 68 as Gln, CCG for residue 75 as Leu, and GAG for residue 76 as Ala

C:Keywords: actin binding; EF hand

F:205 311/Domain: spectrin/dystrophin repeat homology <SP18>

F:320 425/Domain: spectrin/dystrophin repeat homology <SP19>

Query Match 6.0%; Score 99; DB 2; Length 475;

Best Local Similarity 20.4%; Pred No. 3; Matches 64; Conservative 47; Mismatches 135; Indels 68; Gaps 8;

```

Qy 35 PHPGDGLDTHSGLSNSSMTTRELQYWMQNCQPMWVVLFEIASAPITFPF-KVSKF 93
Dh 46 PCGATGVLESFGLKIDNTIGKEITQPLAQPVHMKELQLAANAGQPLESLEYQF 105
Qy 94 V-----VQIIVIGTGSFQNNKAVLERYSDFAKLQKAL 126
Dh 106 VANVEEBAMINEKMTLVASEYGDTLAIOGLLKHEAFETDFTHKDVNDVCTHGOGL 165

```

QY 127 LKTRREIPIVERPRKHTNFA--EEMICEPPALQEVYGLLYAIPCVPPSPREPLDL 194  
 DB 166 LKKNHHEENISKMKGLNKVSDLEKAAQAPYAKDE-----NSAFIQFN 212  
 QY 185 TRPELPEAF-----GCLPAGQYPALELLIPVLPLEQELTAHCPA-----AAVPAICAV 233  
 DB 213 WKAQVEWIGEKENSLKTDYGRDLSSVQTLTKQETFDAGLOAFQOEGIANITALKDQ 272  
 QY 234 LKCRDLDRA-----FAFA-AGEPAICQLQAPEGHPYVAPLDAAVPLAVA 279  
 DB 273 LLAKHIOSKAIEARHASMFPMTQLFANSAATPYKYLEAQSHRPYEDFLFAVAYASA 332  
 QY 280 LKQDFVTIGERLEE 293  
 DB 333 FNSWFEANEDLTD 346

RESULT 11  
 A35715  
 fodrin alpha chain - human  
 N/Alternate names: alpha II spectrin; nonerythroid alpha-spectrin  
 C/Species: Homo sapiens (man)  
 C/Date: 05-Oct-1990 #sequence revision 05-Oct-1990 #text change 13-Aug-1999  
 C/Accession: A35715; A28580; EC382; A61369; G01810; S54877; S65689  
 R/Moon, P.T.; McMahon, A.P.  
 J. Biol. Chem. 265, 4427-4433, 1990  
 A/Title: Generation of diversity in nonerythroid spectrins Multiple polypeptides are pr  
 A/Reference number: A35715; MUID:90170946; PMID:2307671  
 A/Accession: A35715  
 A/Status: preliminary  
 A/Molecule type: mRNA  
 A/Residues: 1-2472 <MOO>  
 A/Cross-references: GR:J05233; NID:9273205; PIDN:AAAS1790.1; PID:9579106  
 R/McMahon, A.P.; Giesbreght, D.H.; Champion, J.E.; Haller, J.A.; Lacey, S.; Carrille, B.,  
 Differentiation 34, 68-78, 1987  
 A/Title: cDNA cloning, sequencing and chromosome mapping of a non-erythroid spectrin, hu  
 A/Reference number: A28580; MUID:87277022; PMID:3038643  
 A/Accession: A28580  
 A/Molecule type: mRNA  
 A/Residues: 676-1594 <MCK>  
 R/McMahon, A.P.; Moon, P.T.  
 Biochem. Soc. Trans. 15, 804-807, 1987  
 A/Title: Structure and evolution of a non-erythroid spectrin, human alpha-fodrin  
 A/Reference number: 152382; MUID:88083947; PMID:3691949  
 A/Accession: 152382  
 A/Status: translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 676-1594, 'P', 'PES'  
 A/Cross-references: GB:M24773; NID:9537330; PIDN:AAAS2468.1; PID:9537331  
 R/Mitcha, S.; Langston, A.; Bonifas, J.M.; Fagerlin Jr., E.H.  
 J. Invest. Dermatol. 97, 383-388, 1991  
 A/Title: Biochemical identification of alpha-fodrin and protein 4.1 in human keratinocy  
 A/Reference number: A61369; MUID:91341201; PMID:1875039  
 A/Accession: A61369  
 A/Molecule type: mRNA  
 A/Residues: 'P', '1595-1610', 'E', '1612-1614', 'P', '1616-1624', 'S', '1626-1638', 'P', '1640', 'E', '1642-16  
 2-1813', 'P', '1915-1820', 'K', '1822-1844', 'V', '1846-1850' <MTU>  
 A/Note: the sequence in Fig. 4 is inconsistent with the sequence in Fig. 3 at several po  
 R/Morrow, J.S.  
 submitted to the EMBL Data Library, May 1995  
 A/Reference number: G08473  
 A/Accession: G08473  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 811-1299, 'I', '1301-1529' <MOR>  
 A/Cross-references: EMBL:U02396; NID:9836668; PIDN:AA60364.1; PID:9836663  
 R/Lundberg, S.; Bjoeck, J.; Loeftvenberg, L.; Backmann, L.  
 submitted to the EMBL Data Library, May 1995  
 A/Description: Cloning and characterization of two putative calcium-binding sites in hum  
 A/Reference number: S54877  
 A/Accession: S54877  
 A/Status: preliminary

A/Molecule type: mRNA  
 A/Residues: 2059-2312, 'P', '2314-2346', 'EF', '2349-2333' <LUT>  
 A/Cross-references: EMBL:X86901  
 R/Lundberg, S.; Bjoeck, J.; Loeftvenberg, L.; Backmann, L.  
 Eur. J. Biochem. 230, 658-665, 1995  
 A/Title: Cloning, expression and characterization of two putative calcium-binding sites  
 A/Reference number: S65689; MUID:95331304; PMID:7607240  
 A/Accession: S65689  
 A/Status: preliminary; nucleic acid sequence not shown  
 A/Molecule type: mRNA  
 A/Residues: 2320-2346, 'EF', '2349-2406' <LU2>  
 A/Cross-references: EMBL:X86901  
 C/Genetics:  
 A/Genes: GDB:SPTANI  
 A/Cross-references: GDB:120385; OMIM:182810  
 A/Map position: 9q34.1-9q34.1  
 C/Superfamily: spectrin alpha chain; calmodulin repeat homology; SH3 homology; spectrin/  
 C/Keywords: actin binding; EF hand  
 F:466-571/Domain: spectrin/dystrophin repeat homology <SP5>  
 F:572-676/Domain: spectrin/dystrophin repeat homology <SP6>  
 F:677-782/Domain: spectrin/dystrophin repeat homology <SP7>  
 F:974-1021/Domain: SH3 homology <SH3>  
 F:2090-2195/Domain: spectrin/dystrophin repeat homology <SP19>  
 F:2284-2310/Domain: spectrin/dystrophin repeat homology <SP20>  
 F:2333-2355/Domain: spectrin/dystrophin repeat homology <SP21>  
 F:2366-2398/Domain: calmodulin repeat homology <EF2>

Query Match 6.0%; Score 99; DB 2; Length 2472;  
 Best Local Similarity 22.6%; Pred. No. 22;  
 Matches 69; Conservative 41; Mismatches 141; Indels 54; Gaps 11;

QY 37 PGPDGLDTHSLSSNSSTTRELQYWNOKCPKHYKLLFEIASAIEEP-KYSKRVV 95  
 DB 1818 PAIQVLDTGKTLSCNNTGKEELQCELAQFEVHEMLKVLAAAGCELESLEYQFVA 1877  
 QY 96 YQIIVYQTSFNNKAVL--ERRYSDFPKLOKALK-----TRFEIEDV----- 138  
 DB 1878 N--VEEFAMINEGMLVASDYGDTLALQGLIKKEAEIAFVYKLVNEYJTHQ 1934  
 QY 139 EPPFKHUTGNFAEEMICEPPALQEVYGLLYAIPCVPPSP-----EFLDFTLPELPEAF 193  
 DB 1935 DLIKRY---NHHEENISKMKGLNKVSDLEKAAQAPYAKDE-----NSAFIQFN 1991  
 QY 194 -----GCLPAGQYPALELLIPVLPLEQELTAHCPA-----AAVPAICAVILC----- 236  
 DB 1992 IGBKENSILKTDYGRDLSSVQTLTKQETFDAGLOAFQOEGIANITALKDQILAAKHVQS 2051  
 QY 237 -----HRDLRPAAPAFAGEPALOR--IQAPEGHPYVAPLDAAVPLAVAGKDFVTLQ 288  
 DB 2052 YALEAPHASMFPMTQLFANSAATPYKYLEAQSHRPYEDFLFAVAYASAFNSFEVAE 2111  
 QY 289 ERLEE 293  
 DB 2112 EDLTD 2116

RESULT 12  
 T31608  
 hypothetical protein Y5088A.d - Caenorhabditis elegans  
 C/Species: Caenorhabditis elegans  
 C/Date: 29 Oct-1999 #sequence revision 29-Oct-1999 #text change 29-Oct-1999  
 C/Accession: T31608  
 R/Steward, C.  
 submitted to the EMBL Data Library, September 1999  
 A/Reference number: Z21047  
 A/Accession: T31608  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-459 <WIL>  
 A/Cross-references: EMBL:AL117200; NID:41549770; PIDN:CAR55047.1; CESP:Y5088A.d  
 A/Experimental source: clone Y5088A  
 C/Genetics:  
 A/Genes: CESP:Y5088A.d

A:Items: 7/3; 63/3; 95/2; 142/1; 169/2; 216/2; 379/3

## Query Match

Best Local Similarity 5.9%; Score 98; DB 2; Length 459;  
C:Species: Homo sapiens (man)  
C:Date: 01-Feb-2002 #sequence\_revision 01-Feb 2002 #text\_change 19-Apr-2002  
Matches 63; Conservative 41; Mismatches 77; Indels 104; Gaps 14;

```

QY 39 PGCHLDTHSGLSNNSMTTRELQO-----YWNQKCGMKVHVKLLFFIASRI 95
DB 212 PTOH-----GLKALDLVTLRELAQKRVNVIPIAKSDTCKELLFK-AKLLELSOKI 265
QY 86 E-----ERKYSKEVVOIIVIGTGSFD-----NKAVALERY 117
DB 266 DVTFTDDELVTSTYKEMVYCPFAVV -GS:DPVYKREJCQVAFQYFWG:VEVEN 321
QY 118 ---SDPAKLOKALLTPPEEIEDEVEPRH--LTGFAEMICPPPAL:CEYGLTAIP 172
DB 322 ESHCHDVKIKFALITNVEMRO---RTHSELVYPRDRLQKIKIDGE-TGPKIIEK 376
QY 173 CVARSEFLDPLTPPE--LPEAFGLPAQGYPPALELLPLVLQCEKLTNCPAAVAPAL 230
DB 377 LAQKHFHODEFRRLTLPEEF-----OKKLD----- 404
QY 231 CAVILTHPLDLPPEAFAGEPALQPLQAEGRHYAPLLDAMVR 275
DB 405 ---VTEGIMRKYEEGLAAPER---EVHENYNEASKLDEIR 440

```

## RESULT 13

Hypothetical protein DPFZ564F052.1 human (fragment)

C:Species: Homo sapiens (man)  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 13-Aug-1999  
Accession: T08691  
R:Rupert-Hoelt, A.; Lamber, J.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
submitted to the Protein Sequence Database, May 1999  
A:Reference number: Z16470  
A:Accession: T08691  
A:Molecule type: mRNA  
A:Residues: 1-420 <DUE>  
A:Cross-references: EMBL, AF049989  
A:Experimental source: fetal brain; clone DPFZ564F052  
C:Function:

A:Notes: DPFZ564F052.1

## Query Match

Best Local Similarity 5.9%; Score 97.5; DB 2; Length 420;  
C:Species: Homo sapiens (man)  
C:Date: 01-Feb-2002 #sequence\_revision 01-Feb 2002 #text\_change 19-Apr-2002  
Matches 39; Conservative 26; Mismatches 51; Indels 21; Gaps 7;

```

QY 90 VSRFVVYQIIV-LGTSPFNNAVALERYSP--ATLQVA---LKTFFETEDVEF 140
DB 79 IETFTYPIITKTSRGEFFSFEFEVRRPYQDFLMKLEEMHPTLIPLPE-----KF 133
QY 141 PRKHITGNFAEEMICERRPALCEYGLTAIPCVSRSPPLDPLTPP-EL-----PEAFG 194
DB 134 IVFTVFPFHDNPTETFEFAKIRFAPITARNPTCTFENEFPTFLAQWELSSHFEGQFG 133
QY 195 CL RAGQYPPALELLR 210
DB 194 LLSKTYQTVKAVASSMK 210

```

## RESULT 14

A59404

plectin [imported] - human  
C:Species: Homo sapiens (man)  
C:Date: 01-Feb-2002 #sequence\_revision 01-Feb 2002 #text\_change 19-Apr-2002  
Accession: C59404, A59404  
R:Jin, C.G.; Maegerle, C.; Castanon, M.J.; Hauptmann, R.; Wiche, G.  
Proc. Natl. Acad. Sci. U.S.A. 93, 4278-4283, 1996  
A:Title: Human plectin: organization of the gene, sequence analysis, and chromosome loca  
A:Reference number: C59404; MIM:602163; PMID 861955  
A:Accession: C59404  
A:Status: preliminary

A:Molecule type: DNA  
A:Residues: 1-4684 <STO>  
A:Cross references: GB:CAA91196, MID:91296662, PIRN:CAA91196.1  
C:Superfamily: plectin; alpha-actinin actin-binding domain homology; ribosomal protein S

## Query Match

Best Local Similarity 5.9%; Score 97; DB 2; Length 4684;  
C:Species: Homo sapiens (man)  
C:Date: 01-Feb-2002 #sequence\_revision 01-Feb 2002 #text\_change 19-Apr-2002  
Matches 78; Conservative 51; Mismatches 109; Indels 146; Gaps 15;

```

QY 44 DTHSGLSNNSMTTRELQOYQWQKCPWKVHVKLLFEIA--SAPIEPKY----- 90
DB 404 TWCQGVAN-----RLQPLWCEVP FLVLLLLQMPHHTAAFEPRPSSFEIEIL 454
QY 91 SKFVYQIIVIGTGSFNNKA-----VIEPPY-----SPKATQKAL 127
DB 455 WQFLFFEMELPEADEKMSKJ:YQSLDQVW:ALGLVFPV:YHPLVEFEW:KFLVVA 514
QY 128 L---KTFPEEIEDEVEPRKHULTNFAEEMICPPPAL:CEYGLTAIPCVSRSPH 179
DB 515 LPERQILRSFEPLFCLOPIVTKLQMERGLTEEDLHDAKLLSLDILILAAQKVPQAGE 574
QY 180 FLDFLTRE--LPEAFGLPAQGYPPALELLPLVLQCEKLTNCPAAVAPAL 230
DB 575 VEPDLDKADSMPLFLFNDVQTLKDPHSGQWYVPVPLHPIVAIPFVYNIPLRAGVA 634
QY 222 CPAAV-----PALCAVLIC 236
DB 635 APATQVAQVTLQSVOPPELFDSTLPYIQLILAWVENGRPVNGAWYVPLPSVEAGLAS 694
QY 233 HRDLDPPEAFAGEPALQPLQAEGRHYAPLLDAMVR 275
DB 695 HPGHNGSTFFPGA KIEPARSDEQSLSPATPAPYPPG:CPILN:QVYVILN:SGVAP:PS 751
QY 270 LDMYRLVAVLCKDFVTQERLEE 293
DB 752 LESLHSFVAATKELMLNEKEEE 775

```

## RESULT 15

C83318

conserved hypothetical protein Pa6613 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
Accession: C83318  
R:Stover, C.K.; Plam, X.Q.; Ewitt, A.L.; Mizoguchi, S.D.; Wanner, P.; Hickey, M.J.; Br  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.R.; Folger, K.R.; Kae, A.; Lathig, R.; Lim,  
J.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathog  
A:Reference number: A82950; MIM:2043737; PMID:10984043  
A:Accession: C83318  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-441 <STO>  
A:Cross-references: DB:AF034696, DB:AF034696.1, MIM:2043737, PIRN:A03741, EMBL:AF034696.1  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: Pa6613  
C:Superfamily: Haemophilus influenzae conserved hypothetical protein HI1590

## Query Match

Best Local Similarity 5.8%; Score 96; DB 2; Length 441;  
C:Species: Pseudomonas aeruginosa (strain PA01)  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
Matches 80; Conservative 47; Mismatches 132; Indels 114; Gaps 16;

```

QY 40 DCHLDTHSGLSNNSMTTRELQOYQWQKCPWKVHVKLLFEIAAPRIFPVSVFVY 96
DB 72 DAHFTISAIVSG---VKEIPQAVEVAKCHAAQVGPFTLLFVNEVHPNKSQDAFLPY 127
QY 97 ---GIVY-CTGSEDNHRAVLER---RYSDFAKLOKAL 127
DB 128 VEFY---FFGATTEFSELMALISFAVYVLLLEALALSTVWALHPIFGVPPIL 197
QY 128 -----LKTFFETEDVEPRKHULTGNFAEEMICERR-----A 160

```



```
Db 188 RLPESFOVLLAARDGGRLLNLLENADLAEDGSETISPELLQNLGDTTRRFRDKGGEA 247
QY 161 LQEVTLGLVYAIRCVFRSP-----EFID---..FLTPPELR---EAFGCLRAQCYP 202
Db 248 FYDQISALH--KSVYRGSNPDALYWFAMMLDGGCDPLYIARRVYRMASEVY---NADP 301
QY 203 PALELLPVLPLQEKLTATACPAAVPAICAVLCPHPTLDPPEAFAPAGCEPALQPLCAPEG 262
Db 302 RALGCLSAMVQERLGSPEGELAVQAIVYLCAPKSNAYYSAFNNAARDVAESGSRE- 360
QY 263 HPVYAP--ILDAMPVL-----AYALGPDFTTICEPIESCIPEPTPR 302
Db 361 ----VPLHLRNAPTKLMKSLGYGEERYAHDEPDNAYAGEDY--FPEDLEPRRYQPVPR 414
QY 303 GITLKELTVREYL 315
Db 415 GLELKIRDKLEHL 427
```

Search completed: April 20, 2003, 14:08:00  
Job time : 51 secs



GenCore version 5.1.4\_dp5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd

CM protein - protein search, using sw model

Run on: April 20, 2003, 13:19:31 ; Search time 24 Seconds

(without alignments) 546 106 Million cell updates/sec

Title: US-09-816-697a-2

Perfect score: 1650

Sequence: 1 MASPRHGPSGCMPTQCT...PPPTPRGCTIKETLVEPEYLH 316

Scoring table: BIOSUM62

Gap 10.0, Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1233.5	76.0	313	1	Q9DZ55 mus musculus
2	413.5	25.1	373	1	SNX1_HUMAN
3	123	7.5	342	1	SNX2_HUMAN
4	118	7.2	169	1	SNXO_HUMAN
5	112.5	6.8	435	1	SNXN_HUMAN
6	111	6.7	2477	1	SPCN_CHICK
7	102.5	6.2	1127	1	WMT1_YEAST
8	102	6.2	2472	1	SPCN_RAT
9	100.5	6.1	343	1	SNXG_HUMAN
10	99	6.0	2472	1	SPCN_HUMAN
11	98.5	6.0	204	1	SNXN_HUMAN
12	98.5	6.0	705	1	SYN1_HUMAN
13	97	5.9	270	1	SNXB_HUMAN
14	97	5.9	4684	1	PLEI1_HUMAN
15	96.5	5.8	387	1	SNX7_HUMAN
16	96	5.8	595	1	SNX9_HUMAN
17	95.5	5.8	339	1	NCF4_MOUSE
18	95.5	5.8	387	1	SNX7_MOUSE
19	92.5	5.6	450	1	SNX4_HUMAN
20	92	5.6	706	1	SYN1_BOVIN
21	92	5.6	3674	1	SPCR_HUMAN
22	91	5.5	1565	1	DNM1_HUMAN
23	90.5	5.5	23	1	SNXM_HUMAN
24	90.5	5.5	344	1	SNXG_RAT
25	90	5.5	1446	1	IE18_PRYKA
26	90	5.5	1461	1	IE18_PRYVF
27	89.5	5.4	3321	1	PN2_HUMAN
28	89.5	5.4	455	1	PPB_CHICK
29	89.5	5.4	670	1	SYN1_MOUSE
30	89	5.1	4687	1	PLEI1_PAT
31	88	5.1	162	1	SNX3_YEAST
32	88	5.1	519	1	SNX2_MOUSE
33	87.5	5.3	1130	1	SBCC_BACCU

34	87	5.3	964	1	PLEI1_MOUSE	Q9QX51 mus musculus
35	86.5	5.2	222	1	Y6E1_DEIRA	Q9Y222 drosophila
36	86.5	5.2	893	1	YW92_CAEEL	P34531 caenorhabditis
37	86.5	5.2	1130	1	YL17_CAEEL	O11102 caenorhabditis
38	86	5.2	499	1	YE29_THFAC	O9H102 thermoplasma
39	86	5.2	519	1	SNX2_HUMAN	O60749 homo sapien
40	86	5.2	642	1	DPX3_SALTY	P74876 salmonella
41	86	5.2	1235	1	RP4A_CHICK	O90640 gallus galli
42	85.5	5.2	455	1	RBB_COTXA	O90643 coturnix co
43	85.5	5.2	455	1	PPB_HUMAN	P10826 homo sapien
44	85.5	5.2	522	1	SNX1_HUMAN	O13596 homo sapien
45	85.5	5.2	522	1	SNX1_PAT	Q90637 rattus norv

## ALIGNMENTS

## RESULT 1

SNXK\_MOUSE STANDARD, PRT, 313 AA.

Q9DZ55, 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Sorting nexin 20.

GN SNX20.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_Taxid=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J;

RA Hong W.;

RT "A new member (SNX20) of the sorting nexin protein family."

RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Cecum;

RA MEDLINE=21085660; PubMed=11217851,

RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,

RA Aizawa K., Hara A., Fukunishi Y., Komori H., Adachi S.,

RA Aizawa K., Iizawa Y., Nishi K., Kiyosawa H., Kasukawa T., Saito R.,

RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,

RA Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,

RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,

RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Persle G., Quackenbush J.,

RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,

RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,

RA Blake J., Roffelli D., Rojunga N., Carninci P., de Bonaldo M.F.,

RA Brownstein M.O., Bull C., Fletcher C., Fujita M., Gariboldi M.,

RA Gustincich S., Hill D., Hofmann W., Hume D.A., Kamiya M., Lee N.H.,

RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,

RA Nordone P., Ring P., Ringwald M., Rodriguez I., Sakamoto N.,

RA Sasaki H., Sato K., Schoenbach C., Seva T., Shibata Y., Storch K.F.,

RA Suzuki H., Taya-Oka K., Wang Y.H., Weitz C., Whitaker C., Wilking L.,

RA Wynshaw-Bois A., Yoshida K., Hasegawa Y., Kawaji H., Kontseki S.,

RA Hayashizaki Y.;

RT "Functional annotation of a full-length mouse cDNA collection."

RL Nature 409:685-690(2001).

CC - FUNCTION: May be involved in several stages of intracellular

CC trafficking (by similarity).

CC - SIMILARITY: BELONGS TO THE SORTING NEXIN FAMILY.

CC - SIMILARITY: CONTAINS 1 PX DOMAIN.

CC - SIMILARITY: CONTAINS 1 PX DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

```

DR EMBL: AF395844; AA073125.1;
DR EMBL: AK018632; BAB31317.1;
DR MGI: 1918857; 9130017C7RLK.
DR InterPro: IPR001683; PX.
DR Pfam: PF00787; PX; 1.
DR SMART: SM00312; PX; 1.
DR PROSITE: PS0195; PX; 1.
DR Transport: Protein transport.
KT DOMAIN 71 188
SQ SEQUENCE 313 AA, 35998 MW, EAPCD04FA22707C3 CRG64,

Query Match 76.0%; Score 1253.5; DB 1; Length 313;
Best Local Similarity 77.5%; Pred. No. 1.6e-96;
Matches 244; Conservative 25; Mismatches 43; Indels 3; Gaps 1;

QY 1 MASPEHSGSPGWMPTTQCTAPTDQEPATGDDLPHPGDDGLDTHSGLSNSSMTTREL 60
DB 1 MASPEHSGSPGWMPTTQCTAPTDQEPATGDDLPHPGDDGLDTHSGLSNSSMTTREL 57

QY 61 QOYWKQKGRWYVTLFPIASAPTEERPVSVFYVYQIIIVITQSGFNNKAVLERYSDF 120
DB 58 QEHMKQKSPWYVTLFPIASAPTEERPVSVFYVYQIIIVITQSGFNNKAVLERYSDF 117

QY 121 AKLOKALITFPEETEDVEFPKHLITGNFAEMICEPPALQEVYGLIYAIRCVPSSEF 180
DB 118 EPIQKALITFPEETEDVEFPKHLITGNFAEMICEPPALQEVYGLIYAIRCVPSSEF 177

QY 181 LDEITPPELPPEAFGCIIPACQYPALELLRVLPLEKLTATGPAVAVALCAVLCHIDL 240
DB 178 LDEITPPELPPEAFGCIIPACQYPALELLRVLPLEKLTATGPAVAVALCAVLCHIDL 237

QY 241 DRPAEFAAGERALQRLQAREGHRYAPALDAMVRLAVALGDFPTLDERLESQLEPPT 300
DB 238 DRPAEFAAGERALQRLQAREGHRYAPALDAMVRLAVALGDFPTLDERLESQLEPPT 297

QY 301 PRGITLTKEITVPEYL 315
DB 298 HRDATLTKEITVPEYL 312

RESULT 2
SNXL HUMAN STANDARD; PRT; 373 AA.
AC Q969T3; Q9BR16;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Sorting nexin 21.
GN SNX21 gp 220pp161;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
CX MGI TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Hong W.;
RT "A new member (SNX21) of the sorting nexin protein family."
RN [2]
RA Submitted (JUN-2001) to the EMBL/Genbank/DBJ databases.
RA Stravides G.S., Huckle E.J., Deloukas P.;
RN [3]
RA Submitted (MAY-2001) to the EMBL/Genbank/DBJ databases.
RP SEQUENCE FROM N.A.
RA MEDLINE: 2173749; PubMed: 1179052;
RA Deloukas P., Matthews L.H., Aspuril J., Burton J., Gilbert J.G.R.,
RA Jones M., Stavrides G., Almeida J.P., Babaei A.K., Bagunley C.L.,
RA Bailey J., Batlow K.E., Bates K.N., Beard L.M., Beare D.M.,
RA Beaskey C.P., Bird C.P., Blakely S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clae C.M.,
RA Cobley S., Cobley V.F., Collier P.F., Connor P.E., Corby N.R.,
RA Coulton A., Coville G.J., Deadman R., Dhani P.D., Dunn M.,

```

```

RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Graham D.V., Griffiths C., Griffiths M.N.D., Gwilliam P., Hall P.E.,
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Humble E., Hunt A.R., Hunt S.E., Jakesch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Levasseur M.H., Leverisha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McConachie L.J., Mclay K., Murray A.A.,
RA Milne S.A., Mistry D., Moore M.P., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel P., Pearce T.A.V., Peck A.I.,
RA Phillimore R.J.C.T., Prathalingam S.P., Plumb P.W., Pansay H.,
RA Rice C.M., Ross M.T., Scott C.F., Sehra H.K., Showkheen P., Sims S.,
RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA Swann P.M., Sycamore N., Taylor P., Teel L., Thomas P.W., Thorpe A.,
RA Tracey A., Tremans A.C., Vandin M., Wall M., Wallis J.M.,
RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilming I., Wray P.W., Hubbard T., Hurfin P.W., Bentley D.F., Beck S.,
RA Rogers J.;
RT "The DNA sequence and comparative analysis of human chromosome 20."
RL Nature 414:865-871(2001)
CC !. FUNCTION: May be involved in several stages of intracellular
CC trafficking (By similarity).
CC !. SIMILARITY: BELONGS TO THE SORTING NEXIN FAMILY.
CC !. SIMILARITY: CONTAINS 1 PX DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF395845; AA073126.1;
DR EMBL: AL592562; CAC39140.1;
DR EMBL: AL008726; CAC36011.1;
DR Genew: HCRG.16154; C200r1161.
DR InterPro: IPR001683; PX.
DR InterPro: IPR001440; TPR.
DR Pfam: PF00515; TPR; 1.
DR Pfam: PF00787; PX; 1.
DR PROSITE: PS0195; PX; 1.
DR Transport: Protein transport.
KT DOMAIN 129 246
FT CONFIDENT 97 99
SQ SEQUENCE 373 AA; 41365 MW; 83BA4752BAEA7B5 CRG64;

Query Match 25.1%; Score 413.5; DB 1; Length 373;
Best Local Similarity 35.4%; Pred. No. 5.8e-27;
Matches 108; Conservative 44; Mismatches 120; Indels 33; Gaps 6;

QY 31 GPD-LP-----HPPGCHLITHSLSNSSMTTRELQVWQNKQWYVHLL 77
DB 81 GPD-LP-----HPPGCHLITHSLSNSSMTTRELQVWQNKQWYVHLL 77

QY 78 FEIASAPTEERPVSVFYVYQIIIVITQSGFNNKAVLERYSDEPAKLOKALTFTPEIED 117
DB 133 FEIVSANVTEERPVSVFYVYQIIIVITQSGFNNKAVLERYSDEPAKLOKALTFTPEIED 192

QY 138 VEFPKHLITGNFAEMICEPPALQEVYGLIYAIRCVPSSEFRLDTPPELPPEAFGCIIP 197
DB 193 VEFPKHLITGNFAEMICEPPALQEVYGLIYAIRCVPSSEFRLDTPPELPPEAFGCIIP 252

QY 198 AGQYPALELLRVLPLEKLTATGPAVAVALCAVLCHIDLCHPDLPPAEFAAGERALOR 256
DB 253 TLVTEFAALAMAMQDGLDTHSGLSNSSMTTRELQVWQNKQWYVHLL 312

QY 257 LGAPGHPYVAPLITAWPLAVALGDFPTLDERLESQLEPPTPRGIT 310
DB 313 LGAPGHPYVAPLITAWPLAVALGDFPTLDERLESQLEPPTPRGIT 367

QY 311 VPEYL 315
DB 368 IREYL 372

```

## RESULT 3

SNXF HUMAN STANDARD; PRT; 342 AA.

AC Q9NR56; Q9NR55; 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Sorting nexin 15.  
 GN SNX15.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 NX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Parathyroid;  
 RX MEDLINE=21265032; PubMed=11085978;  
 RA Phillips S.A., Barr V.A., Haft D.H., Taylor S.T., Haft C.P.;  
 RT "Identification and characterization of snx15, a novel sorting nexin  
 involved in protein trafficking.";  
 RL J. Biol. Chem. 276:5074-5084(2001).  
 CC -1- FUNCTION: MAY BE INVOLVED IN SEVERAL STAGES OF INTRACELLULAR  
 TRAFFICKING. OVEREXPRESSION OF SNX15 DISRUPTS THE NORMAL  
 ENDOSOMES OR THE TGN.  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOPROTEIN; 1 (SHOWN HERE) AND 2/SNX15A,  
 ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED.  
 CC -1- SIMILARITY: BELONGS TO THE SORTING NEXIN FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 PX DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AF175267; AAF89955.1;  
 CC EMBL; AF175268; AAF89956.1;  
 CC Genbank; HGNC:14978; SNX15.  
 CC MIM: 605964;  
 CC InterPro; IPR001683; PX.  
 CC Pfam; PF00787; PX; 1.  
 CC SMART; SM00312; PX; 1.  
 CC PROSITE; PS0195; PX; 1.  
 CC KMW Transport; Protein transport; Alternative splicing.  
 CC FT DOMAIN 1 130 PX  
 CC VARSPLIC 221 308 ECGAAPSPTVAELATNEVESARLDQPMFGCGEEDGE  
 CC GGTPTATVLSQATLITQALRDEKGAVALAALOGYRGDVHL  
 CC LQGVPS -> G (IN ISOPRO 2).  
 CC SEQUENCE 342 AA; 38291 MW; 33F64A79EAF6BDC CFC64;

Query Match 7.5%; Score 123; DB 1; Length 342;  
 Best Local Similarity 24.9%; Pred. No. 0.0055;  
 Matches 60; Conservative 33; Mismatches 166; Indels 42; Gaps 9;

QY 78 FEIASARIEERKVSFVYVQIIVIGTGSFNNKAVLERYSDFAKLOKALKTFRREIEDV 137  
 Db 13 YTVSDPRTHPKGYTRYKYTAQFISKRPEDVKEVVVMKPYSDPRLHDDLVYTPNLRP 72  
 QY 138 VE----FPRKULTGNFAEMTCFPRALQFVLGLLYAIPCVRSREIFDLTPBELRPAF 193  
 Db 73 LEEPAFPAPQVGFGEASVIEERPKGAEDLRTVHLPALNS-----FQLKEFF 123  
 QY 194 GCLPAGQVPALELIL--FVPLPLCEKLTANCPAAVPAALCAVLCIHR---DLRPAE- 245  
 Db 124 ---RGGEVTRPLEVSRDLHILPRPIPT---PRDDPRLSQLPRERKGLBELEVYVDFP 177  
 QY 246 AFAAGERALQRL-----QABEGHPYVAPLLDAMVPLAVALGPDFVTLQEPLEESQLRRP 299

Db 178 PSSPAQEAIDLFCNESTSEASGPARGPLTEALFDPFSPK-----EEGAPSP 228  
 QY 300 T 300  
 Db 229 T 229

## RESULT 4

SNXO HUMAN STANDARD; PRT; 169 AA.

AC Q9Y343; 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Sorting nexin 24 (SBB131).  
 GN SNX24.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 NX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Zhang W., Man T., Cao X.;  
 RT "Hypothetical human protein SBB131.";  
 RL Submitted (MAR-1999) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Hong W.;  
 RT "SBB131 as a new member (SNX24) of the sorting nexin family.";  
 RL Submitted (JUL-2001) to the EMBL/Genbank/DBJ databases.

CC -1- SIMILARITY: BELONGS TO THE SORTING NEXIN FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 PX DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AF139461; AAD32668.1;  
 CC EMBL; AY044655; AAK96769.1;  
 CC EMBL; BC010886; AAH10886.1;  
 CC InterPro; IPR001683; PX.  
 CC Pfam; PF00787; PX; 1.  
 CC SMART; SM00312; PX; 1.  
 CC PROSITE; PS0195; PX; 1.  
 CC KMW Transport; Protein transport.  
 CC FT DOMAIN 1 125 PX  
 CC SEQUENCE 169 AA; 19818 MW; 651B3964493AFESD CFC64;

Query Match 7.2%; Score 118; DB 1; Length 169;  
 Best Local Similarity 32.1%; Pred. No. 0.006;  
 Matches 40; Conservative 22; Mismatches 90; Indels 22; Gaps 7;

QY 80 IASAPTEPRVSK-FVYVQIIVIGTGSFNNKAVLERYSDFAKLOKALKTFRREIEDV 138  
 Db 5 IPSFRYESDLERGVYVFKIIVLMNG---PHFVEKYESEHALHKYIKK---ITP 56  
 QY 139 EPRKULTGNFAEMTCFPRALQFVLGLLYAIRCVRSRRE---FLDFLT---RPELRE 191  
 Db 57 EIPSKAVR-NWPKVLEGRPGQLETLQAV-----ILENEELPKLFLDFLNVRLPSLPK 110  
 QY 192 AFQCG 195

```

Db      111 AASC 114

RESULT 5
ID      SNXR_HUMAN          STANDARD:      PRT: 435 AA.
AC      096L92;
DT      15-JUN-2002 (rel. 41, Created)
DT      15-JUN-2002 (rel. 41, Last sequence update)
DE      15-JUN-2002 (rel. 41, Last annotation update)
DE      Sorting nexin 27.
DE      SNX27.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX      NCBI_TaxID=9606;
RN      (1)
RP      SEQUENCE FROM N.A.
RA      Hong W.;
RT      Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RL      FUNCTION: May be involved in several stages of intracellular
CC      trafficking (by similarity).
CC      - SIMILARITY: BELONGS TO THE SORTING NEXIN FAMILY.
CC      - SIMILARITY: CONTAINS 1 PX DOMAIN.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@sib.ch)
CC      -----
DR      EMBL; A0044866; AAK97797.1; -
DR      InterPro; IPR001683; PX.
DR      InterPro; IPR001683; PX.
DR      Pfam; PF00787; PX; 1.
DR      Pfam; PF00788; RA; 1.
DR      PROSITE; PS50195; PX; 1.
KW      Transport; Protein transport.
FT      DOMAIN 68 176
FT      DOMAIN 180 269
SQ      SEQUENCE 435 AA; 50880 MW; 8F3AD932ACDCC66 CFC64;

Query Match      6.8%; Score 112.5; DB 1; Length 435;
Best Local Similarity 28.5%; Pred No 0.955;
Matches 35; Conservative 22; Mismatches 37; Indels 29; Gaps 5,

```

```

OS      Gallus gallus (chicken).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC      Gallus.
OX      NCBI_TaxID=9031;
RN      (1)
RP      SEQUENCE FROM N.A.
RA      MEDLINE=89093238; PubMed=2910879;
RA      Maenius V.-M., Saraste M., Salven P., Eraama M., Holm M.,
RA      Lehto V.-P.;
RT      "Primary structure of the brain alpha-spectrin.";
RL      J. Cell Biol. 108:79-93(1989).
RN      [2]
RP      REVISIONS.
RA      Maenius V.-M., Saraste M., Salven P., Eraama M., Holm L.,
RA      Lehto V.-P.;
RL      J. Cell Biol. 108:1177-1178(1989).
RN      [3]
RP      SEQUENCE OF 1695-2153 FROM N.A.
RA      MEDLINE=85284928; PubMed=4029118;
RA      Maenius V.-M., Saraste M., Knowles J., Virtanen I., Lehto V.-P.;
RT      "Sequencing of the chicken non-erythroid spectrin cDNA reveals an
RT      internal repetitive structure homologous to the human erythrocyte
RT      spectrin.";
RL      EMBO J. 4:1425-1430(1985).
RN      [4]
RP      X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 965-1025.
RA      MEDLINE=93063297; PubMed=1279434;
RA      Muscicchio A., Noble M., Pauplit R., Wierenga R., Saraste M.;
RT      "Crystal structure of a Src-homology 3 (SH3) domain.";
RL      Nature 359:851-855(1992).
RN      [5]
RP      X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 966-1025.
RA      MEDLINE=98363217; PubMed=3692637;
RA      Martinez J.C., Pisabarro M.T., Serrano L.;
RT      "Obligatory steps in protein folding and the conformational diversity
RT      of the transition state.";
RL      Nat. Struct. Biol. 5:721-729(1998).
RN      [6]
RP      STRUCTURE BY NMR OF 1763-1872.
RA      MEDLINE=98022917; PubMed=9356261;
RA      Pascual J., Ftouni M., Walcher D., Saraste M., Nilges M.;
RT      "Solution structure of the spectrin repeat: a left-handed
RT      antiparallel triple-helical coiled-coil.";
RL      J. Mol. Biol. 273:740-751(1997).
RN      [7]
RP      STRUCTURE BY NMR OF 2320-2403.
RA      MEDLINE=96067123; PubMed=7588621;
RA      Tiave G., Lacombe J.-P., Ftouni M., Saraste M., Pastore A.;
RT      "Molecular mechanism of the calcium-induced conformational change in
RT      the spectrin EF-hands.";
RL      EMBO J. 14:4922-4931(1995).
RN      [8]
RP      FUNCTION: MORPHOLOGICALLY, SPECTRIN-LIKE PROTEINS APPEAR TO BE
CC      RELATED TO SPECTRIN, SHOWING A FLEXIBLE ROD-LIKE STRUCTURE.
CC      THEY CAN BIND ACTIN BUT SEEM TO DIFFER IN THEIR CALMODULIN-
CC      BINDING ACTIVITY. IN NONERYTHROID TISSUES, SPECTRINS, IN
CC      ASSOCIATION WITH SOME OTHER PROTEINS, MAY PLAY AN IMPORTANT
CC      ROLE IN MEMBRANE ORGANIZATION.
CC      - SUBUNIT: LIKE ERYTHROCYTE SPECTRIN, THE SPECTRIN-LIKE PROTEINS
CC      ARE CAPABLE TO FORM DIMERS WHICH CAN FURTHER ASSOCIATE TO
CC      TETRAMERS.
CC      - SIMILARITY: SPECTRIN-LIKE PROTEINS HAVE FIVE DOMAINS:
CC      (1) N-TERMINAL DOMAIN (N),
CC      (2) DOMAIN BETWEEN THE N-TERMINAL AND MIDDLE DOMAIN (NM),
CC      (3) MIDDLE DOMAIN (M),
CC      (4) DOMAIN BETWEEN THE MIDDLE AND C-TERMINAL DOMAIN (MC),
CC      (5) C-TERMINAL DOMAIN (C).
CC      NM AND MC DOMAINS ARE COMPOSED OF TYPICAL SPECTRIN 106 RESIDUES
CC      REPEATS (1-8 FOR NM AND 12-19 FOR MC) AND ARE HOMOLOGOUS TO EACH
CC      OTHER. N, M, AND C DOMAINS ARE COMPOSED OF SEQUENCES THAT DO NOT
CC      FORM TYPICAL SPECTRIN REPEATS.
CC      - SIMILARITY: BELONGS TO THE SPECTRIN FAMILY.
CC      - SIMILARITY: CONTAINS 1 SH3 DOMAIN.

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; X14518; CAA32662.1; ALT SEQ.  
 DR EMBL; X14519; CAA32663.1; ALT SEQ.  
 DR EMBL; X02593; CAB51571.1; ALT SEQ.  
 DR PIR; A10192; STCHA  
 DR PDB; 1SHG; 31-OCT-93.  
 DR PDB; 1AEY; 15-MAY-97.  
 DR PDB; 1AU3; 07-JUL-97.  
 DR PDB; 1TUC; 01-AUG-96.  
 DR PDB; 1TUD; 01-AUG-96.  
 DR PDB; 1BK2; 16-FEB-99.  
 DR InterPro; IPR002048; EF-hand.  
 DR InterPro; IPR001452; SH3.  
 DR InterPro; IPR02017; Spectrin.  
 DR Pfam; PF00018; SH3; 1.  
 DR Pfam; PF00036; ehand; 2.  
 DR PRINTS; PR00435; Spectrin; 23.  
 DR PRODOM; PD000012; EF-hand; 1.  
 DR PRODOM; PD000066; SH3; 1.  
 DR SMART; SM00054; EFh; 2.  
 DR SMART; SM00326; SH3; 1.  
 DR SMART; SM00150; SPEC; 20.  
 DR PROSITE; PS00019; EF\_HAND; 2.  
 DR PROSITE; PS50002; SH3; 1.  
 KM Cytoskeleton; Membrane; Calmodulin binding; Actin-binding;  
 KM Capping protein; Calcium-binding; Repeat; SH3 domain; 3D-structure.  
 FT DOMAIN 1 14 N-TERMINAL DOMAIN.  
 FT REPEAT 15 119 1.  
 FT REPEAT 120 225 2.  
 FT REPEAT 226 331 3.  
 FT REPEAT 332 437 4.  
 FT REPEAT 438 543 5.  
 FT REPEAT 544 648 6.  
 FT REPEAT 649 754 7.  
 FT REPEAT 755 860 8.  
 FT REPEAT 861 966 9.  
 FT DOMAIN 967 1061 MIDDLE DOMAIN.  
 FT DOMAIN 967 1026 SH3.  
 FT DOMAIN 1062 1256 MC DOMAIN  
 FT REPEAT 1062 1167 11.  
 FT REPEAT 1204 1309 12.  
 FT REPEAT 1310 1415 13.  
 FT REPEAT 1416 1521 14.  
 FT REPEAT 1522 1633 15.  
 FT REPEAT 1634 1739 16.  
 FT REPEAT 1740 1845 17.  
 FT REPEAT 1846 1951 18.  
 FT REPEAT 1952 2058 19.  
 FT REPEAT 2059 2171 20.  
 FT REPEAT 2172 2256 21.  
 FT DOMAIN 2257 2477 C-TERMINAL DOMAIN.  
 FT CA\_BIND 2341 2352 EF-HAND 1 (POTENTIAL).  
 FT STRAND 2384 2395 EF-HAND 2 (POTENTIAL).  
 FT STRAND 971 974  
 FT STRAND 978 978  
 FT TURN 983 984  
 FT STRAND 985 985  
 FT STRAND 988 988  
 FT TURN 990 991  
 FT STRAND 993 998  
 FT STRAND 1004 1009  
 FT TURN 1010 1011  
 FT STRAND 1012 1017  
 FT HELIX 1018 1020

FT STRAND 1021 1024  
 SQ SEQUENCE 2477 AA; 285361 MW; AD4C876994E6AB37 CRC64,  
 Query Match 6.7%; Score 111; DB 1; Length 2477;  
 Best local Similarity 21.2%; Pred. No. 0.61; Mismatches 134; Indels 68; Gaps 8;  
 Matches 66; Conservative 44; Mismatches 134; Indels 68; Gaps 8;

QY 37 PPGDGLDTHSGLSGSSNSTTRELQGYWQNCQCPWYHVLLEIASARIEER-EVSKFEV-94  
 DB 1823 PAIQGLDTHSGLSGSSNSTTRELQGYWQNCQCPWYHVLLEIASARIEER-EVSKFEV-94  
 QY 95 -----VQIIVIGSGSPNNKAVLEPPYSPAKIQALL-128  
 DB 1883 NVEEDAMINEMMTIVASSTVETLAATQGLKHEAEETDTVKHEDVNCVANGEDLI-1942  
 QY 129 KTFERRIEDVEFPFKHLTGNTFA--EEMICEPPALQGEVGLIYAIPTCPSPSEPIDELTTP-186  
 DB 1943 KYNNHHVENITAMVKGILKGVSDLEAAQCPAKINR-----NSAFICFTMY-1999  
 QY 187 PELPEAF-----GCLPAGQYPPALELLPVLPLELQELTANCPA-----AAVPAICAVLL-235  
 LB 1990 AFVBSWISSEKENSLEKTEVYGNLSVQTLTKLETFDAGLQAFQEGIANITALKQQL-2049  
 QY 236 C-----HRLDPRPAEFAAGERALQK--LQAREGHRVYAFLLDAWVRLAVALG-281  
 DB 2050 AAKH:QSYA:IEVPHASLWKEPQQLANSAPYKFLLEAGENHFYVEDLFTTAPPAAPFN-2109  
 QY 282 KDFVTLQERLEE-293  
 DB 2110 SWFENAEEDLTD-2121

RESULT 7  
 ID MDM1 YEAST STANDARD; PRT; 1127 AA.  
 AC 001846; Q04196;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Structural protein MDM1.  
 GN MDM1 OR YML104C OR YH8339.15C.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / AB972;  
 RA Devlin K., Skelton J., Churcher C., Barrell B.G., Rajandream M.A.,  
 RA Walsh S.V.  
 RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE OF 633-1127 FROM N.A., FUNCTION, AND SUBCELLULAR LOCATION.  
 RX MEDLINE=92332595, PubMed=1378448,  
 RA McConnell S.J., Yaffe M.P.;  
 RT "Nuclear and mitochondrial inheritance in yeast depends on novel  
 RT cytoplasmic structures defined by the MDM1 protein.";  
 RL J. Cell Biol. 118:385-395(1992).  
 RN [3]  
 RP SEQUENCE OF 874-1127 FROM N.A.  
 RA Stirling C.J.;  
 RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: Essential for mitotic growth. Mediates organelle  
 CC inheritance.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -!- SIMILARITY: TO S.POMBE SPCC16A11.04.  
 CC -!- SIMILARITY: CONTAINS 1 PX DOMAIN.  
 CC -!- SIMILARITY: CONTAINS 1 PXA DOMAIN.  
 CC -!- CAUTION: Pef 2 sequence differs from that shown due to a  
 CC frameshift in position 638.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation.

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

EMBL; X80835; CAA56793.1; ALT\_INIT  
EMBL; 249210; CAA9114.1; ALT\_FRAME  
EMBL; X66371; CAA47014.1; ALT\_FRAME  
EMBL; X65783; CAA46644.1; ALT\_FRAME  
PIR; A4236; A4236.1  
SGD; S0004572; MDM1  
InterPro: IPR001683; PX  
InterPro: IPR003114; PX\_assoc.  
PIfam; PF00787; PX; 1  
SMART; SM00112; PX; 1  
SMART; SM00113; PX; 1  
PROSITE; PS50195; PX; 1  
Coiled coil.  
FT DOMAIN 85 276 PXA.  
FT DOMAIN 705 762 COILED COIL (POTENTIAL).  
FT DOMAIN 782 905 PX.  
FT CONFLICT 645 645 H -> D (IN REF. 2).  
FT CONFLICT 772 772 Q -> R (IN REF. 2).  
FT CONFLICT 899 899 T -> I (IN REF. 2).  
FT CONFLICT 1063 1063 R -> K (IN REF. 2).  
FT CONFLICT 1081 1081 T -> S (IN REF. 2).  
FT CONFLICT 1096 1096 G -> P (IN REF. 2).  
SEQUENCE 1127 AA, 129980 MW, 709177887DPAF12 CRC64,  
Query Match 6.2%; Score 102.5, DB 1, Length 1127,  
Best Local Similarity 26.3%; Pred. No. 12;  
Matches 49; Conservative 30; Mismatches 52; Indels 55; Gaps 11;

42 HLDTHGLSSN-----SSMTTEL--QQTW--QNCQWYVHLF----- 78  
711 HLLKMLTNNMQLKLLKSSAPTLLEKLEMKELKQYMOVENSNLSFRKTIYRSYP 790  
79 -FIASARIEPKVSKFVVVQIVT-----QTSFNNMVALEPSPFALQALLKTF 131  
791 SENSNGCLAE-----ITYIILNIHFNNGVSSMD-----MARRYNEFFELNYLKKNF 839  
132 PE -FIEVEEPK-----HLTGFAEMICERPALQFYGLVLAIPGVSPSEFL 181  
840 RDLMPQLQDL-FPSKVYKMLYHVTTLYE---EPYGLTFYVLRLLSIEICENIIF 896  
182 DELTRP 187  
896 RPLTDP 901

RESULT A  
SPCN RAT  
ID SPCN RAT STANDARD: PRT, 2472 AA.  
AC P16086; P70477; O88663;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Spectrin alpha chain, brain (Spectrin, non-erythroid alpha chain)  
DE (Alpha-II spectrin) (Fodrin alpha chain).  
CN SYTNMI OR SP7A2.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Wistar; TISSUE=Liver;  
RA Kalamaraki P., Gazzotti P.,  
RT "Structural and functional characterization of the calmodulin and  
RT "calpain binding domains of rat liver alphaII spectrin."  
RI Submitted (Aug-1996) to the EMBL/Genbank/DBJ databases.

[2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Skeletal muscle;  
RA Zhu D., Yrsliji J.A., Porter N.C., Panfili W.P., Ruch R.J.;  
RT "Expressional cloning of alpha-fodrin from rat skeletal muscle."  
RI Submitted (Aug-1996) to the EMBL/Genbank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 1292-2321 FROM N.A.  
RC TISSUE=Kidney;  
RA MEDLINE=8937227; PubMed-2753883;  
RX Hong W., Doyle D.;  
RT "Cloning and analysis of cDNA clones for rat kidney alpha-spectrin";  
RJ J. Biol. Chem. 264:12758-12764(1989)  
CC - FUNCTION: FODRIN, WHICH SEEMS TO BE INVOLVED IN SECRETION,  
CC INTERACTS WITH CALMODULIN IN A CALCIUM-DEPENDENT MANNER AND IS  
CC THUS CANDIDATE FOR THE CALCIUM-DEPENDENT MOVEMENT OF THE  
CC CYTOSKELETON AT THE MEMBRANE (by similarity).  
CC - SUBUNIT: LIKE EPYTHROCYTE SPECTRIN, THE SPECTRIN-LIKE PROTEINS  
CC ARE CAPABLE TO FORM DIMERS WHICH CAN FURTHER ASSOCIATE TO  
CC TETRAMERS.  
CC - SIMILARITY: BELONGS TO THE SPECTRIN FAMILY.  
CC - SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.  
CC - SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
CC - SIMILARITY: CONTAINS 23 SPECTRIN REPEATS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

EMBL; X06845; CAA62550.1;  
EMBL; AF084186; AAC33127.1;  
EMBL; J04828; AAA40770.1;  
DR PIR; A32612; A32612.  
DR HSSP; P07751; 1A33.  
DR InterPro: IPR002048; EF-hand.  
DR InterPro: IPR001452; SH3.  
DR InterPro: IPR002017; Spectrin.  
DR Pfam; PF00018; SH3; 1.  
DR Pfam; PF00036; ehand; 2.  
DR Pfam; PF00435; spectrin; 33.  
DR SMART; SM00354; EFh. 2.  
DR SMART; SM00356; SH3. 1.  
DR SMART; SM00150; Spec. 29.  
DR PROSITE; PS00018; EF\_HAND; 2.  
DR PROSITE; PS50002; SH3; 1.  
KW Cytoskeleton; Membrane; Calmodulin-binding; Actin-binding;  
KW Capping protein; Calcium-binding; Repeat; SH3 domain.  
FT REPEAT 10 42 SPECTRIN 1.  
FT REPEAT 44 147 SPECTRIN 2.  
FT REPEAT 149 253 SPECTRIN 3.  
FT REPEAT 255 359 SPECTRIN 4.  
FT REPEAT 361 465 SPECTRIN 5.  
FT REPEAT 467 571 SPECTRIN 6.  
FT REPEAT 573 676 SPECTRIN 7.  
FT REPEAT 678 782 SPECTRIN 8.  
FT REPEAT 784 888 SPECTRIN 9.  
FT REPEAT 890 955 SPECTRIN 10.  
FT REPEAT 967 1026 SPECTRIN 11.  
FT REPEAT 1062 1089 SPECTRIN 12.  
FT REPEAT 1091 1161 SPECTRIN 13.  
FT REPEAT 1209 1231 SPECTRIN 14.  
FT REPEAT 1233 1337 SPECTRIN 15.  
FT REPEAT 1339 1443 SPECTRIN 16.  
FT REPEAT 1445 1549 SPECTRIN 17.  
FT REPEAT 1551 1656 SPECTRIN 18.  
FT REPEAT 1658 1762 SPECTRIN 19.  
FT REPEAT 1764 1868 SPECTRIN 20.  
FT REPEAT 1870 1974 SPECTRIN 21.  
FT REPEAT 1976 2081 SPECTRIN 21.



FT REPEAT 2091 2195 SPECTRIN 22.  
 FT REPEAT 2205 2310 SPECTRIN 22.  
 FT CA\_BIND 2336 2347 EF\_HAND 1 (POTENTIAL).  
 FT CA\_BIND 2379 2390 EF\_HAND 2 (POTENTIAL).  
 FT CONFLICT 1329 1329 D -> Y (IN REF. 3).  
 FT CONFLICT 1514 1514 V -> L (IN REF. 1).  
 FT CONFLICT 1702 1702 L -> A (IN REF. 3).  
 FT CONFLICT 1971 1972 KL -> NV (IN REF. 1).  
 FT CONFLICT 2205 2206 KL -> NV (IN REF. 1).  
 FT CONFLICT 2310 2321 QOICAFRTTGYT -> HADPGQENRPH (IN REF. 1).  
 SQ SEQUENCE 2472 AA; 294635 MW; 08DGF0A2871278A CRC64;

Query Match 5.2%; Score 102; DB 1; Length 2472;  
 Best Local Similarity 20.4%; Pred. No. 3.4;  
 Matches 64; Conservative 46; Mismatches 132; Indels 72; Gaps 8;

QY 37 PCPDGDLTHSGLSNNSMTTRELQOYMONQKRWKHYLLFEIASARIEER-KVSKFV-94  
 DB 1818 PAIGVLDLTKKSLSDNTTIGCEIQQPLAQFVHKWELKQLAANGRLLESLEVCQFVA-1877  
 QY 95 -----VQIIVIQTSFDNNKAVLEPPYSPAKIQKALL-128  
 DB 1878 NVEEEMANINEKMTLVASEDYDGLTAAIQGLKHEAFETFTVHKDYNVCTVQGLI-1937  
 QY 129 KTFPEIPEDEFPFPHLTGNFA--EEMICEPPRALQEVLLVAIPCVRSPEFLDELTR-186  
 DB 1938 KNNHHEENISSKMGKGLKXVSDLEKAAAKAKLDE-----NSAFIQFHWK-1984  
 QY 187 PELPFAF-----GCLPAQYPPALELLLVLPCEKLTANCPA-----AAVPAICAVLL-235  
 DB 1985 AVVESMTGEKENSILKTDYVCPRLSSVQTLTKCEFPAGLQAFQCEGIANTMLKQGLL-2044  
 QY 236 CHRDLDPRAFAFAAGEPALCP-----LQAEGRHYVPLDAMPLAVA-279  
 DB 2045 AAKHIC--SPALFAPHASLMPWTQLANSATPKYKLLAOSHPKVEDLFTPAKASA-2102  
 QY 280 LGKDFVTLQERLEE-293  
 DB 2103 FNSWFEINAEEDLTD-2116

RESULT 9  
 SPCN HUMAN STANDARD; PRT; 343 AA.  
 ID SNXG HUMAN STANDARD; PRT; 343 AA.  
 AC P57768;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Sorting nexin 16.  
 GN SNX16.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OK NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Hanson B.J., Hong W.;  
 RL Submitted (SEP-2000) to the EMBL/Genbank/DBJ databases.  
 CC - FUNCTION: May be involved in several stages of intracellular  
 CC trafficking (By similarity).  
 CC - SIMILARITY: BELONGS TO THE SORTING NEXIN FAMILY.  
 CC - SIMILARITY: CONTAINS 1 PX DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF105779; AAC25676.1; -  
 DR HSRF; P14598; IGDS.

DR InterPro; IPR001683; PX.  
 DR Pfam; PF00787; PX; 1.  
 DR SMART; SM00312; PX; 1.  
 DR PROSITE; PSS0195; PX; 1.  
 KM Transport; Protein transport; Coiled coil.  
 FT DOMAIN 105 218 PX  
 FT DOMAIN 223 278 COILED COIL (POTENTIAL).  
 SQ SEQUENCE 343 AA; 39153 MW; EBE722392A46FF5 CRC64;

Query Match 6.1%; Score 100.5; DB 1; Length 343;  
 Best Local Similarity 27.9%; Pred. No. 0.4;  
 Matches 41; Conservative 29; Mismatches 52; Indels 25; Gaps 8;

QY 47 SGLSSNSMTTRELQOYMON-QKRWKHYLLFEIASARIEERKVFVYQIIV-100  
 DB 78 TGTASSIEVSTRPRDTEQNPETVNMEDRSTPTITIGYEV---MEER--AKFTYKILIV-131  
 QY 101 TGTGTFNNKAVLEPPYSPAKIQ---VALLTTFPEIPEDEFPFPHLTGNFAFEMICP-157  
 DB 132 KKT---PESWVVPFAPYTPSPPLNDYIKEMPPAPPLALP---PPWPKDYNADPLEDP-184  
 QY 158 RRALQEVLLGLVAIPCVRS---REFL-181  
 DB 185 QLGGLQAFLLQVLVAHKDIANCLAVREFL-211

RESULT 10  
 SPCN HUMAN STANDARD; PRT; 2472 AA.  
 ID Q13813; Q16626, Q9FCVC, Q13186;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Spectrin alpha chain, brain (Spectrin, non-erythroid alpha chain)  
 DE (Alpha-II spectrin) (Fodrin alpha chain).  
 GN SPTAN1 OR SPTA2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OK NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Moon R.T., McMahon A.P.;  
 RL MEDLINE=90170948; PubMed=2307671;  
 RT "Generation of diversity in nonerythroid spectrins. Multiple  
 RT polypeptides are predicted by sequence analysis of cDNAs encompassing  
 RT the coding region of human nonerythroid alpha-spectrin.";  
 RL J Biol Chem 274:4437-4443(1999)  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Tissue=Fetal brain;  
 RC Tissue=Fetal brain;  
 RA Cianci C.D., Morrow J.S.;  
 RL "Complete cDNA sequence of human alpha II fetal brain spectrin.";  
 RT Submitted (FEB-1997) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 676-1595 FROM N.A.  
 RC Tissue=Lung;  
 RX MEDLINE=88083942; PubMed=3691949;  
 RA McMahon A.P., Moon R.T.;  
 RT "Structure and evolution of a non-erythroid spectrin, human  
 RT alpha-fodrin.";  
 RL Biochem. Soc. Trans. 15:804-807(1987).  
 RN [4]  
 RP SEQUENCE OF 676-1595 FROM N.A.  
 RX MEDLINE=87277023; PubMed=3038643;  
 RA McMahon A.P., Giebelhaus D.H., Champion J.E., Bailes J.A., Lacey S.,  
 RA Carritt B., Henchman S.K., Moon R.T.;  
 RT "cDNA cloning, sequencing and chromosome mapping of a non-erythroid  
 RT spectrin, human alpha-fodrin.";  
 RL Differentiation 34:68-78(1987).  
 RN [5]  
 RP ERROR.  
 RA McMahon A.P., Giebelhaus D.H., Champion J.E., Bailes J.A., Lacey S.,

RA Carilt B., Henchman S. K., Moon P. T.:  
 RL Differentiation 34:241-241(1987)  
 RN (6)  
 RP SEQUENCE OF 811-1529 FROM N. A., VARIANT 1LE-1300, AND MUTAGENESIS.  
 RC TISSUE=Petal brain;  
 RX MEDLINE=97146462; PubMed-8993318;  
 RA Stach P. P., Cinci C. D., Glantz S. B., Zhang Z., Morrow J. S.,  
 RT "Site-directed mutagenesis of alpha II spectrin at codon 1175  
 RT modulates its mu-calpain susceptibility.";  
 RL Biochemistry 36:57-65(1997).  
 RN (7)  
 RP SEQUENCE OF 1073-1349 FROM N. A.  
 RA Murakami N., Speed W. C., Seaman M. I., Zychowski P. L., Wetzelberg L.,  
 RA Pakesiot A. J., Kidd J. R., Kidd K. K.:  
 RT "Association and linkage analyses of the nonerythroid alpha-spectrin  
 RT (SPPAN1) gene on chromosome 9q34 with a large Swedish kindred.";  
 RL Submitted (May-1999) to the EMBL/GenBank/DBJ databases.  
 CC - FUNCTION: FODRIN, WHICH SEEMS TO BE INVOLVED IN SECRETION,  
 CC INTERACTS WITH CALMODULIN IN A CALCIUM-DEPENDENT MANNER AND IS  
 CC THUS CANDIDATE FOR THE CALCIUM-DEPENDENT MOVEMENT OF THE  
 CC CYTOSKELETON AT THE MEMBRANE.  
 CC - SUBUNIT: LIKE ERYTHROCYTE SPECTRIN, THE SPECTRIN-LIKE PROTEINS  
 CC ARE CAPABLE TO FORM DIMERS WHICH CAN FURTHER ASSOCIATE TO  
 CC TETRAMERS.  
 CC - SIMILARITY: BELONGS TO THE SPECTRIN FAMILY.  
 CC - SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.  
 CC - SIMILARITY: CONTAINS 1 SH3 DOMAIN  
 CC - SIMILARITY: CONTAINS 2 SPECTRIN REPEATS  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 DR EMBL: J05243; AAA51790.1; -  
 DR EMBL: U83867; AA841498.1; -  
 DR EMBL: M24773; AA552468.1; -  
 DR EMBL: M18627; AA551702.1; -  
 DR EMBL: U26386; AA860364.1; -  
 DR EMBL: AF148808; AA626722.1; -  
 DR HSSP: P07751; IAEY.  
 DR GeneW: HGNC:11273, SPTAN1  
 DR MIM: 182810; -  
 DR InterPro: IPR002048; EF hand.  
 DR InterPro: IPR001452; SH3.  
 DR InterPro: IPR002017; Spectrin.  
 DR Pfam: PF00018; SH3; 1.  
 DR Pfam: PF00036; efhand; 4.  
 DR PRINTS: PR00435; spectrin; 32.  
 DR PRODOM: PD000012; EF-hand; 1.  
 DR PRODOM: PD000066; SH3; 1.  
 DR SMART: SM00054; Efh; 2.  
 DR SMART: SM00126; SH3; 1.  
 DR SMART: SM00150; SPEC; 20.  
 DR PROSITE: PS00018; EF HAND; 1.  
 DR PROSITE: PS50002; SH3; 1.  
 RW Cytoskeleton, Membrane, Calmodulin binding, Actin binding,  
 RW Calcium binding, Repeat, SH3 domain, F-lync:phasm.  
 FT REPEAT 10 42 SPECTRIN 1.  
 FT REPEAT 44 147 SPECTRIN 2.  
 FT REPEAT 149 253 SPECTRIN 3.  
 FT REPEAT 255 359 SPECTRIN 4.  
 FT REPEAT 361 465 SPECTRIN 5.  
 FT REPEAT 467 571 SPECTRIN 6.  
 FT REPEAT 573 676 SPECTRIN 7.  
 FT REPEAT 678 782 SPECTRIN 8.  
 FT REPEAT 784 888 SPECTRIN 9.  
 FT REPEAT 890 955 SPECTRIN 10.  
 FT DOMAIN 967 1026 SH3.

FT REPEAT 1062 1089 SPECTRIN 11.  
 FT REPEAT 1091 1161 SPECTRIN 12.  
 FT REPEAT 1208 1231 SPECTRIN 13.  
 FT REPEAT 1233 1337 SPECTRIN 14.  
 FT REPEAT 1339 1343 SPECTRIN 15.  
 FT REPEAT 1445 1449 SPECTRIN 16.  
 FT REPEAT 1451 1549 SPECTRIN 17.  
 FT REPEAT 1551 1656 SPECTRIN 18.  
 FT REPEAT 1658 1762 SPECTRIN 19.  
 FT REPEAT 1764 1868 SPECTRIN 20.  
 FT REPEAT 1870 1974 SPECTRIN 21.  
 FT REPEAT 1976 2081 SPECTRIN 22.  
 FT REPEAT 2091 2195 SPECTRIN 23.  
 FT REPEAT 2305 2310 SPECTRIN 24.  
 FT CA BIND 2336 2347 EF-HAND 1 (POTENTIAL).  
 FT CA BIND 2379 2390 EF-HAND 2 (POTENTIAL).  
 FT SITE 1176 1177 CLEAVAGE (BY MU-CALPAIN).  
 FT VARIANT 1300 1300 T -> I.  
 FT CONFLICT 150 150 /FTID-VAR\_01227.  
 FT CONFLICT 498 498 N -> K (IN REF. 2).  
 FT CONFLICT 737 737 F -> S (IN REF. 2).  
 FT CONFLICT 737 737 V -> I (IN REF. 2).  
 FT CONFLICT 1586 1586 Q -> QLSKL (IN REF. 2).  
 FT CONFLICT 1595 1595 F -> R (IN REF. 3 AND 4).  
 FT CONFLICT 1625 1625 N -> S (IN REF. 2).  
 FT CONFLICT 1671 1671 IA -> FD (IN REF. 2).  
 FT CONFLICT 1918 1918 DA -> D (IN REF. 2).  
 FT CONFLICT 2347 2348 AG -> EF (IN REF. 2).  
 FT CONFLICT 2448 2448 I -> Y (IN REF. 2).  
 SQ SEQUENCE 2472 AA, 284279 MM, CDAL9C4462A7028D CRC64;  
 Query Match 6.0%; Score 99; DB 1; Length 2472;  
 Best Local Similarity 27.6%; Pred No. 6;  
 Matches 69; Conservative 41; Mismatches 141; Indels 54; Gaps 11;  
 QY 37 PGPDSCHLDTGSLSSSSMTPELQYQWQNCQPKWVKVILFETASAPIEPF-KVSEFV 95  
 DB 1818 PAIQGVLDLTQCKKSLPDNTIGKEIQCRLAQFVEHMKELKOLAAARQRLSESLQYQVFA 1877  
 QY 96 VQIIIVIGTGFQDNKAVL--EPRYSDFAKIQKALK-----TFPEIEDV----- 138  
 DB 1878 N---VEEEAWINKEKNTLVASEDYQDTLALIQGLKHEAFETAFVHKRVNVEVCTWQ 1934  
 QY 139 EFPFPHLTGNFAFEMTCEPPALQEVILGLVYVPCVPSRP-----EFLDFLTPELREAF 193  
 DB 1935 DLITKYN NHHEEISKMKGLGQYVDELEKAAK-KAVNDEKQALFVHWKADYVEEM 1991  
 QY 194 -----GCTPAAGYPPALLLVLPVLEKLTAKCPA-----AAVPAVAVLLC----- 246  
 DB 1992 ICEKENSUKTDYGRDLSSVOTLLTKQETPDAGIQAFQCGEIGNITALKQDLAAKHVQS 2051  
 QY 237 -----HPDLPPAPAFACEPALQP--LQAPFGHPYVAPLLDAMPVLAVALGDFVTLQ 288  
 DB 2052 KALEAPHASLKPMSQOLLANSAAKKLLEQSHFRVVDLFTFAKKSASFNSFENAE 2111  
 QY 289 ERLLE 293  
 DB 2112 EDLTD 2116  
 RESULT 11  
 SNXN\_HUMAN  
 AC 096L93; Q9HSU0; STANDARD; FRT; 204 AA.  
 ID 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Sorting nexin 23.  
 GN GNX23.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 NC NCB1\_TaxID=9606;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RA Hong W.;  
 RT "A new member (SNX23) of the sorting nexin family.";  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 RN (2)  
 PP SEQUENCE FROM N.A.  
 RC Tissue-Lung;  
 RA Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,  
 Nakajima Y., Mizuno T., Morinaga M., Tanigami A., Fujiwara T., Ono T.,  
 Yamada K., Fujii Y., Ozaki K., Hirao K., Ohmori Y., Ota T., Suzuki Y.,  
 Obayashi M., Nishi T., Shibahara T., Tanaka T., Nakamura Y.,  
 Isegai T., Sugano S.;  
 RA "NEO human cDNA sequencing project";  
 RT Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: May be involved in several stages of intracellular  
 trafficking (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE SORTING NEXIN FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 PX DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
 CC -----  
 CC EMBL: AY044654; AAK98768.1;  
 DR EMBL: AK026698; BAB15530.1;  
 DR InterPro: IPR001683; PX  
 DR Pfam: PF00787; PX; 1.  
 DR PROSITE: PS0195; PX; 1.  
 DR Transport: Protein transport.  
 FT DOMAIN 69 183 PX.  
 FT CONFLICT 131 131 P -> A (IN REF. 2).  
 SQ SEQUENCE 204 AA; 23746 MW; 2703467E05C7D8F6 CRC64;  
 Query Match 6.0%; Score 98.5; DB 1; Length 204;  
 Best Local Similarity 42.1%; Pred No 0.31;  
 Matches 24; Conservative 10; Mismatches 22; Indels 1; Gaps 1,  
 QY 115 RRYSDPAKQALLKTFREIEDEVEPRKHLTGFAEMICERRRAQEVYGLLYAI 171  
 Db 106 PPSYFPEMHT-ILKLYAEIALAEFFPYKLPGNDEPVIIEPSSHLEKYLDPDFSV 161  
 RESULT 12  
 ID SYN1\_HUMAN STANDARD; PRT; 705 AA.  
 AC P17600; O75825;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Synapsin I (Brain protein 4.1).  
 GN SYN1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxId=9606;  
 RP SEQUENCE FROM N.A.  
 RC Tissue=Brain;  
 RA MEDLINE=90243651; PubMed=2110562;  
 RA Suedhof T.C.;  
 RT "The structure of the human synapsin I gene and protein.";  
 RL J Biol Chem. 265:7849-7852(1990).  
 RN (2)  
 PP SEQUENCE OF 1-125 FROM N.A.  
 RA MEDLINE=90368667; PubMed=2118519;  
 RA Sauerwald A., Haesche C., Oeschwald R., Kilmann M.W.;  
 RT "The 5'-flanking region of the synapsin I gene. A G+C-rich, TATA- and  
 RT CAAT-less, phylogenetically conserved sequence with cell  
 RT type-specific promoter function.";

RL J. Biol. Chem. 265:14932-14937(1990).  
 RN (3)  
 RP SEQUENCE OF 1-258 FROM N.A.  
 RA Grafham D.;  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: NEURONAL PHOSPHOPROTEIN THAT COATS SYNAPTIC VESICLES.  
 CC BINDS TO THE CYTOSKELETON, AND IS BELIEVED TO FUNCTION IN THE  
 CC REGULATION OF NEUROTRANSMITTER RELEASE.  
 CC -1- SUBCELLULAR LOCATION: SYNAPSE.  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS, SYNAPSIN IA (SHOWN HERE) AND  
 CC SYNAPSIN IB: ARE PRODUCED BY ALTERNATIVE SPLICING  
 CC -1- PTM: SUBSTRATE OF AT LEAST FOUR DIFFERENT PROTEIN KINASES. IT IS  
 CC PROBABLE THAT PHOSPHORYLATION PLAYS A ROLE IN THE REGULATION OF  
 CC SYNAPSIN I IN THE NERVE TERMINAL.  
 CC -1- SIMILARITY: BELONGS TO THE SYNAPSIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
 CC -----  
 CC EMBL: M58378; AAC41930.1;  
 DR EMBL: M58321; AAC41930.1; JOINED.  
 DR EMBL: M58341; AAC41930.1; JOINED.  
 DR EMBL: M58351; AAC41930.1; JOINED.  
 DR EMBL: M58353; AAC41930.1; JOINED.  
 DR EMBL: M58359; AAC41930.1; JOINED.  
 DR EMBL: M58371; AAC41930.1; JOINED.  
 DR EMBL: M58372; AAC41930.1; JOINED.  
 DR EMBL: M58373; AAC41930.1; JOINED.  
 DR EMBL: M58374; AAC41930.1; JOINED.  
 DR EMBL: M58375; AAC41930.1; JOINED.  
 DR EMBL: M58376; AAC41930.1; JOINED.  
 DR EMBL: M58377; AAC41930.1; JOINED.  
 DR EMBL: M58378; AAC41931.1; ALT. SEQ.  
 DR EMBL: M58321; AAC41931.1; JOINED.  
 DR EMBL: M58341; AAC41931.1; JOINED.  
 DR EMBL: M58351; AAC41931.1; JOINED.  
 DR EMBL: M58353; AAC41931.1; JOINED.  
 DR EMBL: M58359; AAC41931.1; JOINED.  
 DR EMBL: M58371; AAC41931.1; JOINED.  
 DR EMBL: M58372; AAC41931.1; JOINED.  
 DR EMBL: M58373; AAC41931.1; JOINED.  
 DR EMBL: M58374; AAC41931.1; JOINED.  
 DR EMBL: M58375; AAC41931.1; JOINED.  
 DR EMBL: M58376; AAC41931.1; JOINED.  
 DR EMBL: M58377; AAC41931.1; JOINED.  
 DR EMBL: M55301; AAA60608.1;  
 DR EMBL: AL009172; CAA15657.1;  
 DR PIR: A35363; A35363.  
 DR PIR: B35363; B35363.  
 DR HSSP: P17599; IAUJ.  
 DR Genew: HGNC:11494; SYN1.  
 DR MIM: 313440;  
 DR InterPro: IPR001359; Synapsin.  
 DR Pfam: PF02078; Synapsin; 1.  
 DR Pfam: PF02750; Synapsin C; 1.  
 DR PRINTS: PR01368; SYNAPSIN.  
 DR PROSITE: PS00415; SYNAPSIN 1; 1.  
 DR PROSITE: PS00416; SYNAPSIN 2; 1.  
 KM Synapse; Phosphorylation; Neutone; Repeat; Actin-binding;  
 KM Alternative splicing.  
 FT DOMAIN 1 28 A.  
 FT DOMAIN 29 112 B (LINKER).  
 FT DOMAIN 113 420 C (ACTIN-BINDING AND SYNAPTIC-VESICLE  
 FT BINDING).  
 FT DOMAIN 421 655 D (PRO-RICH LINKER).  
 FT MOD\_RES 656 705 E. PHOSPHORYLATION (BY PKA AND CAMK1).  
 FT MOD\_RES 658 658 PHOSPHORYLATION (BY CAMK2).  
 FT MOD\_RES 568 568





```

FT CONFLICT 706 706 Q -> R (IN REF. 2).
FT CONFLICT 886 886 N -> Y (IN REF. 2).
FT CONFLICT 1002 1002 V -> A (IN REF. 2).
FT CONFLICT 1309 1309 V -> L (IN REF. 2).
FT CONFLICT 1321 1321 V -> V (IN REF. 2).
FT CONFLICT 1334 1334 L -> L (IN REF. 2).
FT CONFLICT 1534 1534 I -> M (IN REF. 2).
FT CONFLICT 1662 1662 A -> T (IN REF. 2).
FT CONFLICT 1688 1688 MLC -> RLR (IN REF. 2).
FT CONFLICT 1767 1767 Q -> E (IN REF. 2).
FT CONFLICT 1789 1789 L -> A (IN REF. 2).
FT CONFLICT 1910 1910 K -> P (IN REF. 2).
FT CONFLICT 2154 2154 K -> N (IN REF. 2).
FT CONFLICT 2160 2160 S -> P (IN REF. 2).
FT CONFLICT 2215 2215 S -> Q (IN REF. 2).
FT CONFLICT 2244 2244 S -> A (IN REF. 2).
FT CONFLICT 3027 3027 K -> E (IN REF. 2).
FT CONFLICT 3310 3310 F -> L (IN REF. 2).
FT CONFLICT 3361 3361 F -> L (IN REF. 2).
FT CONFLICT 3408 3408 F -> L (IN REF. 2).
FT CONFLICT 3447 3447 S -> A (IN REF. 2).
FT CONFLICT 3531 3531 G -> S (IN REF. 2).
FT CONFLICT 3580 3580 P -> S (IN REF. 2).
FT CONFLICT 3589 3589 K -> Q (IN REF. 2).
FT CONFLICT 3596 3596 E -> Q (IN REF. 2).
FT CONFLICT 3616 3616 N -> H (IN REF. 2).
FT CONFLICT 3686 3686 V -> A (IN REF. 2).

```

Query Match 5.9%; Score 97, DB 1; Length 4684;  
 Best Local Similarity 20.3%; Pred. No. 19;  
 Matches 78; Conservative 51; Mismatches 109; Indels 146; Gaps 15;

```

QY 44 DTHSGSSNSMTTRELQOYQWQKCPWVHVLLEPA---SARIEPRV----- 90
DB 404 DVNDGVRAN-----HILPMDYR---ELVILLDMHMHHTAAEERRFSSFEIIL 454
QY 91 SKPVVYQIIVICTGSEFNKA-----VLERRY-.....SDFAKLOAL 127
DB 455 WSOFLKFKEMELPAKPADYRSGVIGVSLGAVQACQLKVRPGYHRLDVEKEMKCLHVAI 514
QY 128 L---KTRFRFIEVFPRKILITGNFAEMICERP-----PAQEVGLLYAIPVPSRF 179
DB 515 LERRKOLRSEFEPRLCQIPIVTKLQMEAGLCEQCHQADALQSDIPLAAGKVPORAGE 574
QY 180 FLDFLTRPE--LREARQ---QUPAQYPRALFELLPLVPLQKETA-----H 221
DB 575 VERDLKXADMIRLLFNVDVQTLKTRPHRGCEGMYPVYVNIHPRIVAIETENVLPKAGVA 634
QY 222 CPAAAV-----PALCAVLLC 236
DB 635 APAQVAQVTLQGVRRPELEDETLRYLQDLAWVEENQHPVDGAFWGVILPSVEACIGS 694
QY 237 HRLIDRIAFAPAGPRALQFLAPFGH-----RYVAPL----- 269
DB 695 HRLIHOSIEFQA---KIEPARSPQSLSPATRGATPRPLSPILNQVAKLNSSTAPLPS 751
QY 270 LDAMVRLAVALGDFVTLQERLEE 293
DB 752 LESLHSTVAATAKELMWLNKEEE 775

```

RESULT 15  
 ID SNX7 HUMAN STANDARD; PRT; 387 AA.  
 AC Q9UNH6; U9Y427; U96FW9;  
 DT 16-OCT-2001 (Ref. 40, GenBank)  
 DT 16-OCT-2001 (Ref. 40, Last sequence update)  
 DT 15-JUN-2002 (Ref. 41, Last annotation update)  
 DR Sorting nexin 7.  
 GN SNX7.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

```

OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=21378165; PubMed=11485546;
RA Teasdale F.D., Lord D., Heugten F., Karlsson L., Gleeson F.A.,
RT "A large family of endosome-localized proteins related to sorting
  nexin 1".
PL Biochem. J. 358:7-16(2001)
RN (2)
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RC Duesterhoeft A., Luyber J., Mewes H. W., Gassenhuber J., Wiemann S.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
CC Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -! FUNCTION: May be involved in several stages of intracellular
  trafficking.
CC -! SIMILARITY: BELONGS TO THE SORTING NEXIN FAMILY.
CC -! SIMILARITY: CONTAINS 1 PX DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation
  at the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@sib.ch).
CC EMBL: AF121857; AAD27830.1;
CC DR EMBL: AL049289; CAP43229.1; ALT_INIT.
CC DR EMBL: EC010349; AAH0349.1;
CC DR Genew, HCHC.14971; SNX7.
CC DR InterPro; IPR001683; PX.
CC DR Pfam; PF00787; PX.1.
CC DR SMART; SM00312; PX.1.
CC DR PROSITE; PS01095; PX.1.
CC KW Transport; Protein transport.
FT DOMAIN 30 151
FT CONFLICT 105 195 M -> I (IN REF. 2).
FT CONFLICT 153 153 S -> P (IN REF. 3).
FT CONFLICT 165 165 M -> T (IN REF. 3).
FT CONFLICT 324 324 E -> G (IN REF. 2).
FT CONFLICT 370 370 F -> L (IN REF. 2).
SQ SEQUENCE 387 AA, 45302 MW, E70B3BEC6013DC1F CRC64;

```

Query Match 5.8%; Score 96.5; DB 1; Length 387;  
 Best Local Similarity 28.5%; Pred. No. 1;  
 Matches 39; Conservative 26; Mismatches 51; Indels 21; Gaps 7;

```

QY 96 VSKPVVYQIIVICTGSEFNKAVALFPRYSN---AVLYQA LKRPFRHIVER 140
DB 46 IETPTVFITITSPGRTUSSEFEVFPVYGFLLKYLEIAHFTLLIPLFE . 100
QY 141 PKXILGNFAEMICEPRPALQEVYGLIYAIPLVPSPEFLDFTLP--EL---REAFG 194
DB 101 IYKWEPEFNDGFETPRPALHFFILNPYACHPTLTNEDGFIILTAAMKLSHPRGSP 160
QY 195 CL-RAGQYPRALELLLR 210
DB 161 LLSRMQTVRAVASSMR 177

```

Search completed: April 20, 2003, 14:05:34  
 Job time : 30 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: April 20, 2003, 13:55:21, Search time 85 Seconds  
(without alignments)

766.010 Million cell updates/sec

Title: US-09-816-697A-2

Sequence: 1 MASEHRSRGCMPTTQCT

PPPTPTTTHKLTPEYLH 316

Scoring table: BLASTX62

Gapop 10.0, Gapext 0.5

Searched: 671580 seqs, 20604715 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL 21:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organelle:\*

9: sp\_phage:\*

10: sp\_plant:\*

11: sp\_rodent:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

15: sp\_virus:\*

16: sp\_bacterioph:\*

17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	264	16.0	184	4 Q8WY78	Q8WY78 homo sapien
2	210	12.7	326	5 Q8WY78	Q8WY78 homo sapien
3	185.5	11.2	295	5 Q8WY78	Q8WY78 homo sapien
4	174.5	10.6	199	4 Q8WY78	Q8WY78 homo sapien
5	122	7.4	152	4 Q8WY78	Q8WY78 homo sapien
6	122	7.4	337	11 Q8WY78	Q8WY78 homo sapien
7	117	7.1	952	10 Q8WY78	Q8WY78 homo sapien
8	116	7.0	1010	3 Q8WY78	Q8WY78 homo sapien
9	115.5	7.0	450	5 Q8WY78	Q8WY78 homo sapien
10	115	7.0	248	16 Q8WY78	Q8WY78 homo sapien
11	114.5	6.9	431	5 Q8WY78	Q8WY78 homo sapien
12	113	6.8	169	11 Q8WY78	Q8WY78 homo sapien
13	111	6.7	473	13 Q8WY78	Q8WY78 homo sapien
14	110.5	6.7	577	11 Q8WY78	Q8WY78 homo sapien
15	109.5	6.6	582	11 Q8WY78	Q8WY78 homo sapien
16	107	6.5	520	5 Q8WY78	Q8WY78 homo sapien

17	103.5	6.3	1006	10 Q8S018	Q8S018 oryza sativ
18	101.5	6.2	450	11 Q91Y02	Q91Y02 mus musculu
19	101	6.1	1325	4 Q9UG16	Q9UG16 homo sapien
20	100.5	6.1	450	4 Q9GCA3	Q9GCA3 homo sapien
21	99.5	6.0	515	4 Q9ER07	Q9ER07 homo sapien
22	99.5	6.0	649	4 Q9KX88	Q9KX88 homo sapien
23	99.5	6.0	1065	10 Q9LWZ8	Q9LWZ8 oryza sativ
24	99	6.0	475	11 Q63363	Q63363 rattus norv
25	99	6.0	1041	10 Q9ASA4	Q9ASA4 oryza sativ
26	98.5	6.0	722	4 Q9BQM1	Q9BQM1 homo sapien
27	98.5	6.0	1764	5 Q9N9U8	Q9N9U8 leishmania
28	98	5.9	461	5 Q9U277	Q9U277 caenorhabdi
29	98	5.9	565	5 Q9HCC3	Q9HCC3 drosophila
30	98	5.9	654	11 Q9CT05	Q9CT05 mus musculu
31	97.5	5.9	476	11 Q9D363	Q9D363 mus musculu
32	97.5	5.9	491	5 Q9V793	Q9V793 drosophila
33	97.5	5.9	1295	10 Q9FNB3	Q9FNB3 arabidopsis
34	97	5.9	395	11 Q9CZK0	Q9CZK0 mus musculu
35	97	5.9	595	11 Q91VH2	Q91VH2 mus musculu
36	97	5.9	1015	10 Q94J02	Q94J02 oryza sativ
37	97	5.9	1030	10 Q9AWS4	Q9AWS4 oryza sativ
38	96.5	5.8	336	4 Q8WVD4	Q8WVD4 homo sapien
39	96	5.8	441	16 Q910M5	Q910M5 pseudomonas
40	96	5.8	1038	10 Q94CM3	Q94CM3 oryza sativ
41	95.5	5.8	373	4 Q8TE00	Q8TE00 homo sapien
42	95.5	5.8	373	16 Q9PA13	Q9PA13 xylella fas
43	95.5	5.8	398	16 P73091	P73091 synchocyst
44	95.5	5.8	580	3 Q94547	Q94547 schizosacch
45	95.5	5.8	1075	10 Q9AWZ3	Q9AWZ3 oryza sativ

## ALIGNMENTS

RESULT 1

ID Q8WY78 PRELIMINARY: PRT: 184 AA.

AC Q8WY78: 01-MAR-2002 (TREMREL: 20, Created)

DT 01-MAR-2002 (TREMREL: 20, Last sequence update)

DT 01-JUN-2002 (TREMREL: 21, Last annotation update)

DE PP3993.

OS Homo sapiens (Human).

OC Eukaryota, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi;

OC Mammalia, Eutheria, Primates, Catarrhini, Homiidae, Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Qin W.X., Zhou X.M., Zhang F.F., Jiang H.G., Huang Y., Wan P.F.,

RA Gu J.R.;

RT "Novel human cDNA clones with function of inhibiting cancer cell

RT growth.";

RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF258570; AAC23773.1; -

DR InterPro: IPR001440; TPR.

DR Pfam: PF00515; TPR.1

SQ SEQUENCE 184 AA; 20612 MW; C3C97FE2A1B19559 CRC64;

Query Match: 16.0%, Score 264, DB 4, Length 184;

Best Local Similarity: 37.3%; Pred. No. 6.4e-15;

Matches: 69; Conservative: 27; Mismatches: 77; Indels: 12; Gaps: 3;

QY 138 VERPPRHGHTGNFAEMTCPPPAALQEVLLVAIPVPPSPFFLQFLPPLPFAFGCLR 197

DB 4 ISPPKRLPRNFTAFITAPSPFAFEQFLSHLQAVPELRHAPLQDFVLPFLRPQSLTC 63

QY 198 AGQYPALELLLEVLPLEOKLT-TAHCPAAVPAALCAVLLCHPRLNPFAFAAGSPALQF 256

DB 64 TGYREALALMAVAMQLOGLQGTGFSGRPLTLIAVACHGFIETFGSAPACAKDL 123

QY 257 LQAREGHRFATPLDMVRLAVAGTFTVTLQEPLEESLQPLFPFGIT-----LKEIT 310

DB 124 LQKSLAPLAPLEAHPVLSWPIG-----LDPQSEAPLQALQENGLTPPTPPSLKEEL 178

```

QY 311 VREYL 315
DB 179 IREVL 183

RESULT 2
SEQUENCE FROM N.A.
ID Q95RT6 PRELIMINARY: PRT: 326 AA
AC Q95RT6;
DT 01-DEC-2001 (TEMBLrel. 19, Created)
DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TEMBLrel. 20, Last annotation update)
DE L012265p.
GN CG3077
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA Stapleton M., Broksstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nuncio J., Pacled J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celinker S.,
RL Submitted (0CT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL, AF061149; AAL28697.1;
DR FlyBase; FBgn0031457; CG3077.
DR InterPro; IPR001683; PX.
DR Pfam; PF00787; PX; 1.
SO SEQUENCE 326 AA; 37253 MW; 6BA3C9C907F1A6C3 CRC64.

Query Match 12.7%; Score 210; DB 5; Length 326,
Best Local Similarity 27.6%; Pred. No. 5,6e-10;
Matches 89; Conservative 38; Mismatches 127; Indels 68; Gaps 12;

QY 15 PITQCTARTQGEAPATPDPDIPRHGPDGHLDT---HSGLSNNSMTTELOQ-YWQN-QK 68
DB 2 PTHAWMAKPIHHPTDGD--PRGPD-ELDSPALEAAALDIPPSDKALQGWERTS 58
QY 69 CPWK-----HVLLPELTASAPL-----EPYKSPVYQVIGTGSF-DNNKAVLEPRY 117
DB 59 AEVKPTTDSVLRFDILAHIMPRDGEVKKIPFVYELIVKQAGATEDTPAVTERPY 118
QY 118 SDFAVITQKALLKTPFPEEIEVFFRPYILTCNFAEKICERPALQSYGLIYALFVPRF3 177
DB 119 TDPPELVLGKPGHAPAMAKYFPAVLWGNFSELTIGERSAFAFLTYVASQMWLRDS 178
QY 178 RFEIDFTPELPEAFAGCLPAQGYPALELLIPLVLPJGK-----LTANC 222
DB 179 EYFLRFQIDELTRACQFLDERENEMALPILNCFPLINKIYMNSPVVILICPLVAAQ 238
QY 223 PAAAVP-----ALCAV-----LLGHRILDPAEAPAGERALQPLQAREGHPYAPILD 271
DB 239 TSSVPPHAAERWALLALSFFETLIDL-----LPLVITPLH 276
QY 272 AMVPLAVALGKDFVTLQERLEE 293
DB 277 TCAHLMWKGQDQKPIITDPLTD 298

RESULT 3
SEQUENCE FROM N.A.
ID Q9VQG1 PRELIMINARY: PRT: 295 AA.
AC Q9VQG1;
DT 01-MAY-2000 (TEMBLrel. 13, Created)
DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
DE CG3077 protein.
GN CG3077.

OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA MEDLINE;20196506; PubMed;10731132;
RA Adams M.D., Celinker S.E., Holt P.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Suton G.G., Wortman J.P., Yandell M.D., Zhang C., Chen L.X.,
RA Randon R.C., Rogers Y.-H.C., Blazer V.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.P., Milos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Plamkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Bencos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Broksstein P., Brotler P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Danke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heilmann T.J., Hernandez J.P., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Inegam C.,
RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Kelcham K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Matrei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.W., Murphy K., Murphy L., Muzny D.V., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacled M.,
RA Palazolo K., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reibert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue R.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong P., Sun E.,
RA Switsch R., Tector C., Turner P., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.K., Weisenbach U.,
RA Williams S.M., Woodage T., Worley K.C., Wu P., Yang S., Yao Q.A.,
RA Ye J., Yeh R.F., Yavert J.S., Zhan M., Zhang G., Zhao G., Zheng L.,
RA Zheng X.H., Zhuo F.N., Chung W., Chou X., Chu S., Chu X., Smith H.O.,
PA Gibbs P.A., Myers R.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL, AE003581; AAHS1213.1;
DR FlyBase; FBgn0031457; CG3077.
DR InterPro; IPR001683; PX.
DR Pfam; PF00787; PX; 1.
DR SMART, SMO0312; PX; 1.
SO SEQUENCE 295 AA; 33532 MW; 7D7805C001B0F31D CRC64;

Query Match 11.2%; Score 185.5; DB 5; Length 295,
Best Local Similarity 27.7%; Pred. No. 6.2e-08,
Matches 80; Conservative 38; Mismatches 102; Indels 69; Gaps 13;

QY 37 PGPDGHLDT---HSGLSNNSMTTELOQ-YWQN-QKCPWK-----HVLLPELTASAPL 85
DB 16 PGPD-ELDSPALEAAALDIPPSDKALQGWERTSAEYKPTTDSVLRFDILAH 74
QY 86 -----EERKSKFVVYQIVIGTGSF-DNNKAVLEPRYSPFAKLQKALLKTPFEEIEDVE 139
DB 75 MPDGEDVYIKFVYVYELTVKQPGATHTTQPAKTERPYTTPPELVLGKPGHAPAMAKY 134
QY 140 FPKKILTNFAFEMICEPPPALQEVYGLIYALFVCPVSPPEFLD---FLTPPELPEAFGC 195
DB 135 FPAKYLMAZFSELTIGERSAFAFLTYVASQMWLRDS-----TACQFLDERENEMALPILNCFPL 184
QY 196 LPAQYQVPALELLIPLVLPJGKLTANCPAAAVP-----ALCAV-----LLGHRILDPA 244
DB 185 LNKITYMNSPVVILIC-----PLVAACTSSVPVPHAAEPWALLALSFFETLIDL----- 236

```



CY 245 EAFAGGAFALCPLOARESHRYAPILDMWRIYVALGPDPTLCPLEE 293  
 Db 237 -----LPLYLPLHTCAHLWQPGQDQPPITDPLND 267

## RESULT 4

OSMUR6 PRELIMINARY; PRT; 199 AA.  
 ID OSMUR6  
 AC OSMUR6 (TREMBlrel. 20, Created)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Similar to chromosome 20 open reading frame 161.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=KIDNEY;  
 RA Strausberg R.;  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC019823; AAH19823.1; -  
 DR InterPro; IPR001683; PX; 1.  
 DR Pfam; PF00787; PX; 1.  
 SQ SEQUENCE 199 AA; 21872 MW; D46D7CD4B0D671CC CRC64;

Query Match 10.6%; Score 174.5; DB 4; Length 199;  
 Best Local Similarity 33.1%; Pred. No. 33e-07;  
 Matches 42; Conservative 19; Mismatches 45; Indels 21; Gaps 3;

OY 31 GPD-LP-----HPGPDGHLDTHSGLSSNSTTRELQYMGNCCKMKHVELL 77  
 Db 81 CPDQLPLGDCGSGEAEPSPPPGQW-----GSQLARQLQDFWKSXNTLAPQRL 132  
 OY 78 FEIASARIEERKVKFVVYQIIVICTGSFNNKAVLEPPYSDPAFLKLTPEEIED 137  
 Db 133 FEVTSANVVKPPKPYVYVLTITVIGPPPPCQPAQISPPYCDPELPHNIQPCFFGPMAA 192  
 OY 138 VEFPRKH 144  
 Db 193 ISFPOSH 199

## RESULT 5

ID O14612 PRELIMINARY; PRT; 152 AA.  
 AC O14612;  
 DT 01-JAN-1998 (TREMBlrel. 05, Created)  
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE Hypothetical 17.8 kDa protein.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97397562; PubMed=9253601;  
 RA Gurev S C, Agarwal S K, Manickam P, Olufemi S E, Crabtree J S, Weisemann J M, Kester M, Kim Y S, Emmert-Buck M R, Liotta L A, Spiegel A M, Boguski M, Roe R A, Collins F S, Burns A L, Marx S J, Chandrasekharappa S C;  
 RA "A transcript map for the 2.8-Mb region containing the multiple RT endocrine neoplasia type 1 locus."  
 RL Genome Res 7:72-735(1997).  
 DR EMBL; AF001435; AAB81205.1; -  
 DR InterPro; IPR001683; PX;  
 DR Pfam; PF00787; PX; 1.  
 DR SMART; SM00312; PX; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 152 AA; 17784 MW; 2B3AF6C5A777B8EF CRC64;

Query Match 7.4%; Score 122; DB 4; Length 152;  
 Best Local Similarity 27.5%; Pred. No. 0.0075;  
 Matches 39; Conservative 24; Mismatches 61; Indels 18; Gaps 4;

OY 78 FEIASARIEERKVKFVVYQIIVICTGSFNNKAVLEPPYSDPAFLKLTPEEIED 137  
 Db 133 VTSDFRTHPKGYTEKVAQFISYFRTPEVVEVWVPEYSGFELHDDLAYTHENLEER 72  
 OY 138 VE----FPRKHLTGNAEEMICERPPALQEVGLVYAIPCVPPSPFFLFTPELPEAF 193  
 Db 73 LEEFPAPFPAQVGRFEASVIEERPKGAEELLRFVTHIPALNNS-----POLKEEF 123  
 OY 194 GCLRACQYPALELL--LRVLP 213  
 Db 124 --RGGEVTRPPEVSRLDILP 142

## RESULT 6

ID O91WE1 PRELIMINARY; PRT; 337 AA.  
 AC O91WE1;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Hypothetical 37.7 kDa protein.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=EYE, AND RETINA;  
 RA Strausberg R.;  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC016091; AAH16091.1; -  
 DR MGD; MGI:1916274; 1500032B08RIK.  
 DR InterPro; IPR001683; PX;  
 DR Pfam; PF00787; PX; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 337 AA; 37742 MW; 743923E18ED9EF4B CRC64;

Query Match 7.4%; Score 122; DB 11; Length 337;  
 Best Local Similarity 27.5%; Pred. No. 0.021;  
 Matches 39; Conservative 24; Mismatches 61; Indels 18; Gaps 4;

OY 78 FEIASARIEERKVKFVVYQIIVICTGSFNNKAVLEPPYSDPAFLKLTPEEIED 137  
 Db 133 VTSDFRTHPKGYTEKVAQFISKDPEDIEVWVKYSDFRKLHNDLAYTHRNLEER 72  
 OY 138 VE----FPRKHLTGNAEEMICERPPALQEVGLVYAIPCVPPSPFFLFTPELPEAF 193  
 Db 73 LEEFPAPFPAQVGRFEASVIEERPKGAEELLRFVTHIPALNNS-----POLKEEF 123  
 OY 194 GCLRACQYPALELL--LRVLP 213  
 Db 124 --RGGEVTRPPEVSRLDILP 142

## RESULT 7

ID O9XIM1 PRELIMINARY; PRT; 952 AA.  
 AC O9XIM1;  
 DT 01-NOV-1999 (TREMBlrel. 12, Created)  
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE At2g15900 protein.  
 GN AT2G15900.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eustosids II; Brassicales; Brassicaceae; Arabidopsis.

NCBI\_TaxID=3702;

PN [1] SEQUENCE FROM N.A.

RP STRAIN=CV. COLUMBIA;

PC MEDLINR-20083487; PubMed10617197;

PA Lin X., Kail S., Ponsley S.D., Shea T.P., Benito M., Town C.D.,

RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,

RA Huell C.P., Kerchum K.A., Lee J.J., Penning C.M., Koo H., Moffat K.S.,

RA Cronin L.A., Shen M., Vanaken S.E., Imayam U., Tallon L.J., Gill C.E.,

RA Adams M.D., Carreira A.J., Traas T.H., Gooden H.M., Scmerville C.P.,

RA Copenhagen G.P., Proust D., Nieman W.C., White C., Fisen D.A.,

RA Salzman S.L., Fraser C.M., Venter J.C.;

RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis

thaliana.";

RL Nature 402:761-768(1999).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;

RA Lin X.;

PL Submitted (MAP-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AC006438; ADD1976.1;

DR InterPro: IPR001683, EX.

DR InterPro: IPR001114; PX: 1.

DR Pfam: PF00787; PX: 1.

DR Pfam: PF02194; PXA: 1.

DR SMART: SM00312; PX: 1.

DR SMART: SM00313; PXA: 1.

SO SEQUENCE 952 AA, 10805 MW, 15425.90 Da; R44898 CFC64,

Query Match 7.1%; Score 117; DB 10; Length 952;

Best Local Similarity 26.3%; Pred. No. 0.2;

Matches 41; Conservative 25; Mismatches 76; Indels 20; Gaps 5,

DB 42 H.DHCHGLSSMSMTRELDQVWQKQKMKIVKLLFELASRIERKRVSVYQIIV 101

DB 503 HNDHTICDHCSTIVLHFRGQHLKCR-----VLGAYFFGQSEFVAYSIATV 553

QY 102 GTCGPRHKAVALFRRYSADAKLQKALKTFFETR--DVEFPRKHLTNFAEMICEER- 158

DB 554 DV---ENKTFVFRYSNFERLHRL-----KRIPNVNLQLPKRIFFSSSTEDAFVHPPG 695

QY 159 KALQEVILGLYALRCVPRSPFPLDPTPELDEAPG 194

DB 606 TQDKRYLQDLCTANVAEQHFVPLSLASKVSVSG 641

RESULT 9

Q95SRK3 PRELIMINARY; PRT; 1010 AA.

ID Q95SRK3;

AC Q95SRK3;

DT 01-MAY-2000 (TREMblrel. 13, Created)

DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)

DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)

DE Hypothetical 117.0 kDa protein ClGall 04 in chromosome III.

GN STCC16A11.04.

OS Schizosaccharomyces pombe (fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;

OC Schizosaccharomyces.

OX NCBI\_TaxID=4896;

RP [1]

RP SEQUENCE FROM N.A.

RC STRAIN=972;

RA McDonald R.C., Rajandream M.A., Barrell B.G., Zimmermann W.,

RA Wambolt P.;

PL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.

CC -1. SIMILARITY: TO YEAST MDML.

DR EMBL: AL109957; CARS1076.1;

DR InterPro: IPR001683; PX.

DR InterPro: IPR003114; PX.

DR InterPro: IPR000142; Pogl Gpotein.

DR Pfam: PF00787; PX: 1.

DR Pfam: PF02194; PXA: 1.

Pfam: PF00615; RGS; 1.

DR SMART: SM00312; PX; 1.

DR SMART: SM00313; PXA: 1.

DR SMART: SM00315; RGS; 1.

DR PROSITE: PS0132; RGS; 1.

PW Hypothetical Protein, Transmembrane, Coiled coil, Glycoprotein.

FT TRANSMEM 6 26

FT DOMAIN 567 574

FT DMMIN 671 698

FT CARBOHYD 217 217

FT CARBOHYD 247 247

FT CARBOHYD 323 323

FT CARBOHYD 412 412

FT CARBOHYD 482 482

FT CARBOHYD 541 541

FT CARBOHYD 597 597

FT CARBOHYD 812 812

FT CARBOHYD 840 840

FT CARBOHYD 844 844

FT CARBOHYD 851 851

FT CARBOHYD 980 980

SO SEQUENCE 1010 AA, 11632 MW, 24803.94 Da; I77FIDB CFC64,

Query Match 7.0%; Score 116; DB 3; Length 1010;

Best Local Similarity 27.2%; Pred. No. 0.27;

Matches 44; Conservative 35; Mismatches 65; Indels 18; Gaps 7;

DB 681 PQPDYMSQSE--DSKLPSPRSVSDSEKISKENTPPFAVYTPYEPHGHVPSG 735

QY 110 KAVLFPRYSDFAKLQKALKTFFETFEVPRPHLGNFAEMICPPALQEVGLLY 169

DB 736 MWV ARRFREFELHKKQKQY--DQVSIKFTKSTTSLNKVLLEPRALAEVQLSLF 793

QY 170 AIRCVAKRSREPLDPTPEELR EAFQCLRAQ QYRALEEL 208

DB 794 RMIVVDSKMLFMPISQNTIATQWPHVPRVYVWCTLEVL 835

RESULT 9

Q95SRK3 PRELIMINARY; PRT; 450 AA.

ID Q95SRK3;

AC Q95SRK3;

DT 01-DEC-2001 (TREMblrel. 19, Created)

DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)

DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)

DE LD23236P.

GN CG8726.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

RN [1]

RN SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

RA Stapleton M., Brockstein P., Hong L., Agbayani A., Carlson J.,

RA Champs M., Chaver C., Doyse V., Farfan P., Fries F., George R.,

RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,

RA Nuno J., Paclab J., Paragas V., Park S., Phouanavong S., Wan K.,

RA Yu C., Lewis S.E., Rubin G.M., Celinker S.;

PL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL: AY061314; AAL28862.1;

DR Flybase: FBgn033244, CG8726.

DR InterPro: IPR001683; Euk\_kinase.

DR InterPro: IPR001683; PX.

DR Pfam: PF00369; Pkinase; 1.

DR Pfam: PF00787; PX; 1.

DR Prodom: PD000001; Euk\_kinase; 1.

DR PROSITE: PS0011; PROTEIN\_KINASE\_DOM; 1.

DR ATP-binding; Transferase.

SO SEQUENCE 450 AA, 50748 MW, 11848.90 Da; D1B8FGEFE27E15 CFC64;

RESULT 11

ID	Q9M488	PRELIMINARY:	PRT:	431 AA.
AC	Q9M488			
DT	01-MAY-2000 (TREMBLrel. 13			Created)
DT	01-MAY-2000 (TREMBLrel. 13			Last sequence update)
DT	01-JUN-2002 (TREMBLrel. 21			Last annotation update)
DE	CG3138 protein.			
GN	CG3138.			
OC	Drosophila melanogaster (Fruit Fly).			
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha.			
OC	Ephydroidea; Drosophilidae; Drosophila.			
NCBI_taxonomy=7227;				
SEQUENCE FROM N.A.				
STRAIN=BERKELEY;				
MEDLINE=20196006; PubMed=10731132;				
Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,				
Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,				
George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,				
Sutton G.G., Wortman J.R., Yeates M.D., Zhang Q., Chen D.X.,				
Bratton R.C., Rogers Y.-H.C., Blazey R.G., Chapple M., Pfeiffer B.D.,				
Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.P., McKiss G.L.G.,				
Abdel J.F., Agabran A., An H.-J., Andrews-Plankoch C., Baldwin D.,				
Bailey R.M., Basu A., Bakendale J., Bayraktaroglu T., Beasley E.M.,				
Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bottler P.,				
Borova D., Botchan M.R., Bouck J., Cadieu E., Center A., Chandra I.,				
Burtis K.C., Busam D.A., Butler H., Cadden E., Center A., Chandra I.,				
Cherry J.M., Cawley S., Dahlke S., Deng Z., Mays A.D., Dew I., Dietz S.M.,				
de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,				
Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,				
Durbin K.U., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,				
Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,				
Gioe A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,				
Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,				
Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,				
Jalali M., Kalish F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,				
Kimmel B.E., Kodala C.D., Krat C., Kravitz S., Kulp D., Lai Z.,				
Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,				
Lin X., Matrei B., McIntosh T.C., McLeod M.P., Meckerson D.,				
Meukow G., Mishina N.V., Moberly C., Morris C., Moshnell A.,				
Mouton S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,				
Nelson D.R., Nelson K.A., Nixon K., Nusskern D.P., Pacleb J.M.,				
Palazzo R., Peltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,				
Reinert K., Remington K., Simpson M., Skupski M.P., Smith T.,				
Shue B.C., Siden-Klamos I., Stapleton M., Strong R., Sun E.,				
Spier E.R., Spradling A.C., Turner R., Venter E., Wang A.H., Wang X.,				
Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,				
Wang Z.-Y., Wasserman D.A., Weinstein G.W., Weissbach J.,				
Williams S.M., Woodard T., Worley K.C., Wu D., Yang S., Yao Q.A.,				
Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhu G., Zhu Q., Zheng L.,				
Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhou S., Smith H.O.,				
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;				
"The genome sequence of <i>Drosophila melanogaster</i> ."				
Science 287:2185-2195(2000).				
EMBL, AE003435; AAF46070.1; -				
HSSP, Q12923; 3PZD.				
FLYbase: FBgn0029794; CG3138.				
InterPro: IPR001683; PX.				
InterPro: IPR00159; RA_domain.				
Pfam: PF00787; PX; 1.				
SMART, SM00312; PX; 1.				
SEQUENCE 431 AA; 49535 MW; 9E5F03709DA2F697 CRC64;				
Query Match 6.9%; Score 114.5; DB 5; Length 431;				
Best local Similarity 25.9%; Pred No. 0.12;				
Matches 44; Conservative 20; Mismatches 63; Indels 43; Gaps				

23 TOOEAPATGFDLPHPRGDGLDTHSGLSNNSMTTRELQCYWQNCXKPRMNVHVLLEPIAS 82

DB 41 TOGADRLERQ-----EDGSGSYIDSKSLP-----ISTPD 74

QY 83 APTFEKRVKVVVYVIVVQTSFENNNAVIFPPYSDPAKLYALKTPEEIEDEFFR 142

DB 75 YGVIVNGRGRAYVFNHMA-----GPOUCSPRYPFNFNLSIL-----PREFSGNFPK 123

QY 143 KHLTGNF-----AEMICEPPALQCYGLVIAIRCVRSRREFLDLTPRE 186

DB 124 LPPGWPPQI.SRQQLTPPPGLQCYLEKVCAYVIAESDAVDLTLTTE 171

RESULT 12

Q9CRB0 PRELIMINARY: PRT: 169 AA.

AC Q9CRB0: PRELIMINARY: PRT: 169 AA.

DT 01-JUN-2001 (TRENBLREL. 17, Created)

DT 01-JUN-2001 (TRENBLREL. 17, last sequence update)

DE 5730433116RIK protein (2810011K15RIK protein).

GN 5730433116RIK OR 2810011K15RIK

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NCBI\_TaxID=10090;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=EMBRIO;

RA MEDLINE:21085660; PubMed:11217851;

RA Kawai J, Shinagawa A, Shibata K, Yoshino M, Itoh M, Ishii Y, Aizawa T, Hata A, Furukoshi Y, Konno H, Aichi J, Fukuda S, Aizawa K, Izawa M, Nishi K, Kiyosawa H, Kondo S, Yamanaoka I, Saito T, Okazaki Y, Gotohori T, Bono H, Kasukawa T, Saito P, Kadota K, Matsuda H, Ashburner M, Ratliff S, Casavant T, Fleischmann W, Gaasterland T, Gissi C, King B, Kochava H, Kuehl P, Lewis S, Matsuo Y, Nikaido I, Pesole G, Quackenbush J, Schmitt L, Staudt F, Suzuki R, Tamura K, Wagner L, Wasth T, Sakai K, Oikawa T, Furuno M, Aono H, Baldarelli R, Barth G, Hlake J, Hoffelt D, Hojunga N, Carninci P, de Bonaldo M, Brownstein M, Bult C, Fletcher C, Fujita M, Gariboldi M, Gustincich S, Hill D, Hofmann M, Hume D, Kaita M, Lee N, Lyons P, Marchionni L, Mashima T, Mazzaroli J, Mombauts P, Nordene F, Ring B, Ringwald M, Rodriguez I, Sakamoto N, Sasaki H, Sato K, Schoenbach C, Seya T, Shibata Y, Storch K, Suzuki H, Toyooka K, Wang Y, H, Weitz C, Whitaker C, Williams L, Wyszynski B, Yoshida K, Hasegawa Y, Kawai H, Kohetsuka S, Hayashizaki Y.

RA "Functional annotation of a full-length mouse cDNA collection."

RT Nature 409:685-690 (2001).

RL EMBL: AK017610; BAB30838.1;

DR EMBL: AK017611; BAB28427.1;

DR MGI: MGI-1916476; 2810011K15RIK.

DR MGI: MGI-1922957; 271041115RIK.

DR InterPro: IPR001643; PX.

DR Pfam: PF00787; PX: 1.

DR SMART: SM00112; PX: 1.

SO SEQUENCE 169 AA, 1965 MW, 6764.6D4C272C79 CRC64.

Query Match 6.8%; Score 113; DB 11; Length 169;

Best Local Similarity 31.5%; Pred No 0 051;

Matches 39; Conservative 23; Mismatches 40; Indels 22; Gaps 7.

QY 80 IASARLHRRVSK-FVVYQIVVQTSFNNNAVIFPPYSDPAKLYALKTPEEIEDEFFR 138

DB 5 IFSFHEHSDIIFRGYVFKLEVMNG-----PQHFVEKPESEHAKLKXG---IKTP 56

QY 139 EFPKHLTGNFAEEMICEPPALQCYGLVIAIRCVRSRREFLDLTPRE 191

DB 57 EIRSKIVR-NWVRKVLEFORQGLTLYLQAV-----ILENEELPKLFDPLNVRHLSLPK 110

QY 192 AFGC 195

DB 111 AFGC 114

RESULT 13

ID Q90990 PRELIMINARY: PRT: 473 AA.

AC Q90990:

DT 01-NOV-1996 (TRENBLREL. 01, Created)

DT 01-NOV-1996 (TRENBLREL. 01, last sequence update)

DE 01-DEC-2001 (TRENBLREL. 19, last annotation update)

DE Nonerythroid alpha-spectrin mRNA (fragment).

OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

NCBI\_TaxID=9031;

RN (1)

RP SEQUENCE FROM N.A.

RC MEDLINE=8529823; PubMed 380089;

RA Birkenmeier C S, Podina P M, Pevsny E A, Helfman D M, Hughes S H, Barker J E;

RA "Renaturation homology among the internal repeats of erythroid and nonerythroid spectrin."

RT Proc Natl Acad Sci U S A 82:5671-5676 (1985)

RL EMBL: M11053; AAA49674.1, -

DR HSRF: P07751; IAD3.

DR InterPro: IPR002017; Spectrin.

DR Pfam: PF00435; Spectrin; 5.

DR SMART: SM00150; SPEC; 4.

FT NON\_TER 1

FT NON\_TER 473

FT NON\_TER 473

SO SEQUENCE 473 AA, 53734 MW, 88DEDE9506493984B CRC64;

Query Match 6.7%; Score 111; DB 13; Length 473;

Best Local Similarity 21.2%; Pred No 0 28;

Matches 66; Conservative 44; Mismatches 134; Indels 69; Gaps 8;

QY 37 EPTFHLTHTSTSSNSMTTPVQYVWQYCPWYVLLFEIASRIEER EYSEFV 94

DB 143 PAIQGLDTQKSLSDNTDKREIQGLPAPFVDHMKELKLAAPQGLRESLEYQFVA 202

QY 95

DB 203 NVEEEEMINEMKTVASEDYGLTALQSLLEHNEAFETGTVHNVVNDVANDSD 262

QY 129 KTFPEIEVEVFPFHILGNFA EEMICEPPALQCYGLVIAIRCVRSRREFLDLTPRE 186

DB 264 YNNHNVENTITAYWGLGKGVSDLEKAAQPKATLDE-----NSAPFGNNW 309

QY 187 FELREAF-----GLKAAQYFRALELLLVPLQEKLTANCPA-----AAVPAICAVLL 235

DB 310 ADVESWISQENKNSLNTDQYSPVSSVYTLTWQETFAAGVAFQFJANITATRIQL 359

QY 236 C-----HFDLRFVAFVAAFEFALCF LQAFQHRVYATLAAWRIAAVAG 281

DB 370 AAHTGSAIEFPIASTPPWAGLAAVSAFPRYLLEAGCEHFPVENFLTPAVKACATP 429

QY 282 KDFVTLQERLEE 293

DB 430 SWFENAEEDLTD 441

RESULT 14

ID Q941X3 PRELIMINARY: PRT: 577 AA.

AC Q941X3:

DT 01-DEC-2001 (TRENBLREL. 19, Created)

DT 01-DEC-2001 (TRENBLREL. 19, last sequence update)

DT 01-MAR-2002 (TRENBLREL. 20, last annotation update)

DE B1088C09.3 protein.

GN B1088C09.3.

OS Oryza sativa (Rice).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Euphorbiales; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzoae; Oryza.

NCBI\_TaxID=4530;  
 RN [1] SEQUENCE FROM N A  
 RP STRAIN=CV, NIPONBARE;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa niponbare (GA3) genomic DNA, chromosome 1, BAC  
 clone: B1088C09."  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL, AP003734, BAB6097.1;  
 DR InterPro: IPR001064; CysteAllin.  
 DR InterPro: IPR001683; PX.  
 DR Pfam: PF00787; PX; 1.  
 DR PROSITE: PS00225; CRYSTALLIN BETAGAMMA; UNKNOWN\_1  
 SO SEQUENCE 577 AA; 6286 MW; A3EF68B68FCB31D CRC64;

Query Match 6.7%; Score 110.5; DB 10; Length 577;  
 Best Local Similarity 24.5%; Pred. No. 0.39;  
 Matches 69; Conservative 28; Mismatches 110; Indels 75; Gaps 11;

QY 2 ASPEHPSGCMGRTCT-----ARTQEPATGPDLPHGP----- 39  
 DB 65 SSPPPG-PAMVGADESSAVDAGEGHDLDSSSASTDPLHPPSPSSPT 122  
 QY 40 -----DGHLDHSGLSNSMTTFLQYQWQKCFWKHVKLFEIASARIEER 88  
 DB 123 AIAHDARIEEDGEDDSAPHVPSASDEAPFEVQITVSEPKHAEPAGAGY----- 175  
 QY 89 KVSKEVYQIVIQTSFDNNKAVLERYSDFAKLOKALTKFR-----EEIDVE 139  
 DB 176 -IPSGSYFSYLTTPADGILFPVPPFPFVVALAPLAAYPLFVPAPDSIVECQ 234  
 QY 140 FPPKHLTNPAEEMICEPPALQEVILYAIPCVPRSRFFLDTPRELPAFGC-LPA 198  
 DB 215 VMQRH-----EFVNCPPAAQPYLPPLAHPTIGRSALHDLFEPS....GIPTSA 282  
 QY 199 GQYPPALELLRVLPLQEKLTACPMVAAPALCAVLLCHRD 240  
 DB 283 GESPRS-----DPALSAAMSAAVTAPTAPAKPRDI 314

## RESULT 15

Q91WB6 PRELIMINARY; PRT; 582 AA.  
 AC Q91WB6;  
 DT 01-DEC-2001 (Tremblrel. 19, Created)  
 DT 01-DEC-2001 (Tremblrel. 19, last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, last annotation update)  
 DE Similar to hypothetical protein FLJ20335.  
 OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 CX NCBI\_TaxID=10090;  
 RN [1] SEQUENCE FROM N.A.  
 RP TISSUE=SALIVARY GLAND;  
 RA Strausberg R.;  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL, BC016131, AAH16131.1;  
 DR InterPro: IPR000719; Euk\_Pkinase.  
 DR InterPro: IPR001683; PX.  
 DR InterPro: IPR003124; WH2.  
 DR Pfam: PF00069; Pkinase; 1.  
 DR Pfam: PF00787; PX; 1.  
 DR Pfam: PF02205; WH2; 1.  
 DR ProDom: PD000001; Euk\_Pkinase; 1.  
 DR PROSITE: PS0011; PROTEIN\_TINASE\_DOM; 1.  
 KW ATP-binding; Hypothetical protein; Transferase.  
 SO SEQUENCE 582 AA; 65201 MW; ED5496B61F2E9AA CRC64;

Query Match 6.6%; Score 109.5; DB 11; Length 582;  
 Best Local Similarity 36.4%; Pred. No. 0.48;  
 Matches 32; Conservative 10; Mismatches 39; Indels 7; Gaps 2;

QY 98 IIVIQTSFDNNKAVLERYSDFAPKLOKALTKFREIEIDVEFPPKHLTGNFAEEMICER 157  
 DB 37 IIRVQGIASENSWQIVRYSDFDLNLSLQITGL....SLFLPPKKLIGNMDREFFIAER 22  
 QY 158 RPLAQEYLGLVAIPCVPRSS---PEFLD 182  
 DB 93 QRGQNYLVIMANHVLSNCELKKPLD 120

Search completed: April 20, 2003, 14:07:05  
 Job time : 89 secs

2  
.

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2003, 13:16:06, Search time 39 seconds  
(without alignments)  
1079.672 Million cell updates/sec

Title: US-09-816-697A-2

Sequence: 1 MASPERHPSGCMPTTCT... PPTPTGTCTKLTPTVYH 316

Scoring table:

BLAST/MSM2  
Gap 10 0, Gapext 0 5

Search: 908470 seqs, 13350620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 100%  
Maximum Match 100%

Listing first 45 summaries

Database:

1: A\_Geneseq\_101002:\*  
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1650	100.0	316	22	AA079225
2	1650	100.0	314	22	AA079129
3	464	28.1	148	21	AA03067
4	324	19.6	156	22	AA069008
5	264	16.0	184	23	AA017131
6	255	15.5	580	22	AB027894
7	249	15.1	93	22	AA043390
8	185.5	11.2	295	22	AP059625
9	125.5	7.6	212	22	AB033125
10	118	7.2	173	22	AA017487

11	114.5	6.9	431	22	AB059662	Drosophila melanog
12	112.5	6.8	168	21	AA045187	Human secreted pro
13	112.5	6.8	541	22	AB081568	Human cytoskeleton
14	108.5	6.6	141	22	AA017481	Novel signal trans
15	107.5	6.5	166	22	AB032895	Novel human 313nc
16	107	6.5	520	22	AB061499	Drosophila melanog
17	103.5	6.3	245	22	AB020248	Novel human 313nc
18	103.5	6.3	320	22	AB020248	Novel human 313nc
19	100.5	6.1	344	23	AA023382	Human intracellular
20	100.5	6.1	577	22	AA023382	Human PPG-C-M5.2 p
21	100.5	6.1	578	23	AB097328	Novel human protei
22	99.5	6.0	156	22	AB020890	Novel human diagno
23	99.5	6.0	649	22	AB056680	Novel protein kina
24	99.5	6.0	705	22	AA039328	Human polypeptide
25	99.5	6.0	714	22	AA041114	Human polypeptide
26	99	6.0	141	22	AA017470	Novel signal trans
27	98.5	6.0	66	22	AA017617	Novel signal trans
28	98.5	6.0	165	22	AA017438	Novel signal trans
29	98	5.9	278	22	AA017486	Novel signal trans
30	98	5.9	565	22	AB020665	Drosophila melanog
31	97.5	5.9	491	22	AB060691	Human vesicle traf
32	97	5.9	270	22	AA004763	Human protein sequ
33	97	5.9	270	22	AA035594	Human TNF receptor
34	96.5	5.8	447	22	AA074174	Human sorting nexi
35	96.5	5.8	455	22	AB011613	Human polypeptide
36	96.5	5.8	512	22	AA040526	Human protein sequ
37	96	5.8	595	21	AA043303	Human novel cytol
38	96	5.8	1191	22	AA067940	Human TNF receptor
39	95.5	5.8	259	22	AA068540	Human TNF receptor
40	95.5	5.8	784	22	AA000685	Human TNF receptor
41	95.5	5.8	813	22	AA020684	Human TNF receptor
42	95.5	5.8	1911	22	AA067401	Human acid sequen
43	95.5	5.8	2392	22	AA084758	Human Blip protein
44	95.5	5.8	2392	22	AA084759	Kinase defective B
45	95.5	5.8	2930	22	AA081495	Human Act 2 polype

#### ALIGNMENTS

RESULT 1	AA079225	standard, Protein: 316 AA.
ID	AA079225	
XX	AA079225	
AC	AA079225	
XX	AA079225	
DT	03-JAN-2002 (first entry)	
DE	Amino acid sequence of a human PSGL-1 binding protein..	
XX		
KW	Human; P-selectin glycoprotein ligand binding protein;	
KW	PSGL-1 binding protein; selectin ligand interactor cytoplasmic-1 protein;	
KW	SLIC-1 protein; signal transduction; cytoskeletal organization;	
KW	immune response; inflammatory response; cell adhesion; cell migration;	
KW	cell activation; cell growth; cell differentiation; cell proliferation;	
KW	immune system disorder; cardiovascular disorder; hematopoietic disorder;	
XX	thrombotic disorder.	
OS		
XX	Homo sapiens.	
PN	WO200173028-A2.	
XX		
PD	04-OCT-2001.	
XX		
PF	23-MAR-2001, 2001MO-US09469.	
XX		
PR	24-MAR-2000, 2000US-192104P.	
XX		
PA	(GENE) GENETICS INST INC.	
XX		
PI	Lorenz M, Kriz R, Weich N, Shaw GD;	
XX		
DR	WPI, 2001.616502/71.	

DR N-PSDB; AA165812.  
 XX  
 PT Isolated polynucleotides (SLIC-1) which encode novel P-selectin  
 PT glycoprotein ligand (PSGL-1) binding protein, useful as targets for  
 PT developing modulating agents to regulate a variety of cellular  
 PT processes including signal transduction -  
 XX  
 PS Claim 14, Fig 1, 108pp; English.  
 XX  
 CC The present sequence represents a human P-selectin glycoprotein ligand  
 CC (PSGL-1) binding protein. The protein is a member of the selectin  
 CC ligand interactor cytoplasmic (SLIC-1) family. The SLIC-1 polynucleotides  
 CC and polypeptides are useful as targets for developing modulating agents  
 CC to regulate a variety of cellular processes such as signal transduction,  
 CC cytoskeletal organization, immune and inflammatory responses, inter- and  
 CC intra-cellular communication, adhesion, migration, cell activation,  
 CC growth, differentiation and proliferation. The SLIC-1 proteins provide  
 CC novel diagnostic targets and therapeutic agents to control or modulate  
 CC SLIC-1 molecule-associated disorders such as an inflammatory or immune  
 CC system disorder, a cardiovascular disorder, a cellular proliferation,  
 CC activation, adhesion, growth, differentiation or migration disorder or  
 CC a haematopoietic or thrombotic disorder.  
 CC  
 XX  
 S0 Sequence 316 AA;  
 Query Match 100.0%; Score 1650; DB 22, Length 316,  
 Best Local Similarity 100.0%; Pred. No. 2,3e-163;  
 Matches 316; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MASHEHNGSGCGMPTTQCTARTQOEAATGPDLPHPGPDGHLTHSGLSNSMTREL 60  
 DB 1 MASHEHNGSGCGMPTTQCTARTQOEAATGPDLPHPGPDGHLTHSGLSNSMTREL 60  
 QY 61 CQYWNQKCRKRWKVKILFEITASAPTEEPKVSFVYVQIIVIGTSFNNKAVLEPRYSDF 120  
 DB 61 CQYWNQKCRKRWKVKILFEITASAPTEEPKVSFVYVQIIVIGTSFNNKAVLEPRYSDF 120  
 QY 121 AKIQKALLKTFPEEIEDVEFPFKHLTGNFAEEMICEPPALQFYLSGLVAIPCVPRSPREF 180  
 DB 121 AKIQKALLKTFPEEIEDVEFPFKHLTGNFAEEMICEPPALQFYLSGLVAIPCVPRSPREF 180  
 QY 181 LDFLTPELPEAFAGGCIIPAGQYPPALELLLPVLPQEKLTANCPAAAVPALCAVLLCHRD 240  
 DB 181 LDFLTPELPEAFAGGCIIPAGQYPPALELLLPVLPQEKLTANCPAAAVPALCAVLLCHRD 240  
 QY 241 DRPAEAFAGGPAALQIPAGQYPPALDPAVPLAVALGKDFVTLQEPLEESQLRPT 300  
 DB 241 DRPAEAFAGGPAALQIPAGQYPPALDPAVPLAVALGKDFVTLQEPLEESQLRPT 300  
 QY 301 PRGITLKELTIVREYLH 316  
 DB 301 PRGITLKELTIVREYLH 316  
 RESULT 2  
 AAB43067  
 ID AAB43067 standard; Protein; 334 AA.  
 XX  
 AC AAB43067;  
 XX  
 DT 03 JAN 2002 (first entry)  
 XX  
 DE Amino acid sequence of IBD1prox protein.  
 XX  
 KW Human; inflammatory bowel disease 1 protein, IBD1, IBD1prox,  
 KW intestinal inflammatory disease; apoptosis, NF-kappa B, cancer;  
 KW inflammatory disease; immune disease; cryptogenetic inflammation;  
 KW hemorrhagic rectocolitis; Crohn's disease; Blau syndrome.  
 XX  
 OS Homo sapiens.  
 XX  
 FN FR28067.19-A1.  
 XX

PD 28-SEP-2001.  
 XX  
 PF 27-MAR 2000; 2000FR-0203812.  
 XX  
 PP 27-MAR 2000; 2000FP-0003812.  
 XX  
 PA (DAUS-) FOND DAUSSET-CEPH JEAN.  
 XX  
 PI Hugot JP, Thomas G, Zouali M, Lesage S, Chamaillard M;  
 XX  
 DR WPI; 2001-608364/70.  
 XX  
 DR N-PSDB; AA165593, AA165594.  
 XX  
 PT New human nucleic acids associated with intestinal inflammatory  
 PT disease, useful for diagnosis, prognosis and control of these diseases,  
 PT also related proteins -  
 XX  
 PS Claim 4; Page 73-74; 97pp; French.  
 XX  
 CC The present sequence represents a human protein designated IBD1prox.  
 CC The IBD1prox protein is in proximity to a gene encoding inflammatory  
 CC bowel disease 1 (IBD1) protein, which is associated with intestinal  
 CC inflammatory disease. The IBD1 gene is probably involved in regulation  
 CC of apoptosis and activation of NF-kappa B. The IBD1 and IBD1prox  
 CC polynucleotides are useful as source of probes and primers, as  
 CC source of (anti)sense oligonucleotides, for recombinant production  
 CC of polypeptides, and in screening for interactive compounds. The  
 CC polypeptides are used to raise specific antibodies which useful for  
 CC diagnostic detection or purification of IBD1 and IBD1prox, to screen  
 CC for specific binding agents, potential therapeutic agents. The IBD1 and  
 CC IBD1prox polynucleotides and polypeptides are useful for treatment  
 CC and prevention of inflammatory and/or immune diseases or cancer, where  
 CC associated with mutations in genes corresponding to IBD1 and IBD1prox,  
 CC especially cryptogenetic inflammation of the intestines (hemorrhagic  
 CC rectocolitis, Crohn's disease and Blau syndrome).  
 CC  
 XX  
 S0 Sequence 334 AA;  
 Query Match 100.0%; Score 1650; DB 22, Length 334,  
 Best Local Similarity 100.0%; Pred. No. 2,5e-163;  
 Matches 316; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MASHEHNGSGCGMPTTQCTARTQOEAATGPDLPHPGPDGHLTHSGLSNSMTREL 60  
 DB 1 MASHEHNGSGCGMPTTQCTARTQOEAATGPDLPHPGPDGHLTHSGLSNSMTREL 60  
 QY 19 MASHEHNGSGCGMPTTQCTARTQOEAATGPDLPHPGPDGHLTHSGLSNSMTREL 78  
 DB 19 MASHEHNGSGCGMPTTQCTARTQOEAATGPDLPHPGPDGHLTHSGLSNSMTREL 78  
 QY 61 CQYWNQKCRKRWKVKILFEITASAPTEEPKVSFVYVQIIVIGTSFNNKAVLEPRYSDF 120  
 DB 61 CQYWNQKCRKRWKVKILFEITASAPTEEPKVSFVYVQIIVIGTSFNNKAVLEPRYSDF 120  
 QY 79 CQYWNQKCRKRWKVKILFEITASAPTEEPKVSFVYVQIIVIGTSFNNKAVLEPRYSDF 138  
 DB 79 CQYWNQKCRKRWKVKILFEITASAPTEEPKVSFVYVQIIVIGTSFNNKAVLEPRYSDF 138  
 QY 121 AKIQKALLKTFPEEIEDVEFPFKHLTGNFAEEMICEPPALQFYLSGLVAIPCVPRSPREF 180  
 DB 121 AKIQKALLKTFPEEIEDVEFPFKHLTGNFAEEMICEPPALQFYLSGLVAIPCVPRSPREF 180  
 QY 139 AKIQKALLKTFPEEIEDVEFPFKHLTGNFAEEMICEPPALQFYLSGLVAIPCVPRSPREF 198  
 DB 139 AKIQKALLKTFPEEIEDVEFPFKHLTGNFAEEMICEPPALQFYLSGLVAIPCVPRSPREF 198  
 QY 241 DRPAEAFAGGPAALQIPAGQYPPALDPAVPLAVALGKDFVTLQEPLEESQLRPT 300  
 DB 241 DRPAEAFAGGPAALQIPAGQYPPALDPAVPLAVALGKDFVTLQEPLEESQLRPT 300  
 QY 301 PRGITLKELTIVREYLH 316  
 DB 301 PRGITLKELTIVREYLH 316  
 RESULT 3  
 AAB43067  
 ID AAB43067 standard; Protein; 148 AA.  
 XX  
 AC AAB43067;  
 XX



08-FEB-2001 (first entry)  
Human ORFX QRF2831 polypeptide sequence SEQ ID NO:5662.  
Human; open reading frame; ORFX; detection; cytostatic; hepatocytic;  
vulnary; antiparasitic; antiparkinsonian; neurotropic; neuroprotective;  
anticonvulsant; osteopathic; antidiabetic; immunosuppressant; cardiac;  
immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
hypotensive; dermatological; immunosuppressive; antineoplastic;  
antiviral; antibacterial; antifungal; antineoplastic; antihypertensive;  
antidiabetic; gene therapy; cancer; proliferative disorder; hypertension;  
neurodegenerative disorder; osteoarthritis; graft vs host disease;  
cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
cholesterol ester storage; systemic lupus erythematosus; infection;  
severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
bone damage; cartilage damage; antiinflammatory disease; coagulation;  
thrombosis; contraceptive.  
Homo sapiens.  
WO200058473-A2.  
05-OCT-2000.  
31-MAR-2000; 2000WO-US08621.  
31-MAR-1999; 99US-0127607.  
02-APR-1999; 99US-0127636.  
05-APR-1999; 99US-0127728.  
30-MAR-2000; 2000US-0540763  
(CUPA-) CIPAGEN COPP  
Shinkets RA, Leach M;  
WPI; 2000-602362/57.  
N-PSDB; AAC77276.  
Novel nucleic acids and peptides derived from open reading frame X,  
useful for treating e.g. cancers, proliferative disorders,  
neurodegenerative disorders and cardiovascular disease -  
Claim 11; Page 4838-4839; 5507pp; English.  
AAC74446 to AAC77606 encode the proteins given in AAB40737 to AAB43397,  
which represent the human ORFX open reading frames 1 to 3161. The ORFX  
sequences have activities such as: cytostatic; hepatocytic; vulnary;  
antiparasitic; antiparkinsonian; neurotropic; neuroprotective;  
osteopathic; anticonvulsant; antidiabetic; immunosuppressant;  
immunostimulant; cardiac; thrombolytic; coagulant; vasotropic;  
antidiabetic; hypotensive; dermatological; immunosuppressive;  
antineoplastic; antibacterial; antiviral; antifungal; antineoplastic;  
antihypertensive; antidiabetic. The sequences can be used for determining  
the presence of or predisposition to, or preventing or treating  
pathological conditions associated with an ORFX-associated disorder. The  
nucleic acids can be used to express ORFX proteins in gene therapy  
vectors. The proteins and nucleic acids may be used to treat cancers,  
proliferative disorders, neurodegenerative disorders, osteoarthritis,  
graft vs host disease, cardiovascular disease, diabetes mellitus,  
hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
nocturnal haemoglobinuria, antiinflammatory disease, to enhance  
coagulation; to inhibit thrombosis; and as a contraceptive.  
Sequence 148 AA;  
Query Match 29.1%, Score 464, DB 21, Length 148;  
Best Local Similarity 95.5%, Pred. No 4 4e-40;  
Matches 85; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

7 PGSPGCMPTTCTARTQGEAPATGPDLPHPGFEHLTHSLSSNSMTTRELQVWQN 66  
DB 26 PWEPCMGCPITCTARTQGEAPATGPDLPHPGFEHLTHSLSSNSMTTRELQVWQN 85  
QY 67 QKCRWGVKLFELIASARLEPKVSKFV 95  
DB 86 QKCPWVHFLFELIASARLEPKVSKFV 114  
RESULT 4  
AAM89008  
ID AAM89008 standard; Protein; 156 AA.  
AC AAM89008;  
XX 07-NOV-2001 (first entry)  
DE Human immune/haematopoietic antigen SEQ ID NO:16601.  
XX  
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
XX cytostatic; gene therapy; vaccine; metastasis.  
XX  
XX Homo sapiens.  
XX WO200157182-A2.  
XX 09-AUG-2001.  
XX 17-JAN 2001; 2001WO-US01354.  
XX 31-JAN-2000; 2000US-0179065.  
XX 04-FEB-2000; 2000US-0180628.  
XX 24-FEB-2000; 2000US-0184664.  
XX 02-MAR-2000; 2000US-0186350.  
XX 16-MAR-2000; 2000US-0189874.  
XX 17-MAR-2000; 2000US-0190076.  
XX 18-APR-2000; 2000US-0198123.  
XX 19-MAY-2000; 2000US-0205515.  
XX 07-JUN-2000; 2000US-0209467.  
XX 28-JUN-2000; 2000US-0214886.  
XX 30-JUN-2000; 2000US-0215135.  
XX 07-JUL-2000; 2000US-0216647.  
XX 11-JUL-2000; 2000US-0218880.  
XX 14-AUG-2000; 2000US-0217487.  
XX 11-JUL-2000; 2000US-0217496.  
XX 14-JUL-2000; 2000US-0218290.  
XX 26-JUL-2000; 2000US-0220963.  
XX 14-AUG-2000; 2000US-0220964.  
XX 14-AUG-2000; 2000US-0224518.  
XX 14-AUG-2000; 2000US-0224519.  
XX 14-AUG-2000; 2000US-0225213.  
XX 14-AUG-2000; 2000US-0225214.  
XX 14-AUG-2000; 2000US-0225266.  
XX 14-AUG-2000; 2000US-0225267.  
XX 14-AUG-2000; 2000US-0225268.  
XX 14-AUG-2000; 2000US-0225270.  
XX 14-AUG-2000; 2000US-0225447.  
XX 14-AUG-2000; 2000US-0225757.  
XX 14-AUG-2000; 2000US-0225758.  
XX 14-AUG-2000; 2000US-0225759.  
XX 18-AUG-2000; 2000US-0226279.  
XX 22-AUG-2000; 2000US-0226681.  
XX 22-AUG-2000; 2000US-0226686.  
XX 22-AUG-2000; 2000US-0227182.  
XX 23-AUG-2000; 2000US-0227009.  
XX 30-AUG-2000; 2000US-0228924.  
XX 01-SEP-2000; 2000US-0229287.  
XX 01-SEP-2000; 2000US-0229343.  
XX 01-SEP-2000; 2000US-0229344.  
XX 01-SEP-2000; 2000US-0229345.  
XX 05-SEP-2000; 2000US-0229509.  
XX 05-SEP-2000; 2000US-0229513.  
XX 06-SEP-2000; 2000US-0230437.



```

KW Human: Cancer cell growth inhibitor; cancer; cytostatic.
XX
OS Unidentified.
XX
FN CN1324819-A.
XX
PD 05-DEC-2001.
XX
PF 18-MAY-2000; 2000CN-0115744.
XX
PR 18-MAY-2000; 2000CN-0115744.
XX
FA (SHAN-) SHANGHAI CITY INST ONCOLOGY.
XX
PI Gu J, Yang S;
XX
DR WPI; 2002 281646/33.
XX
DR N-PSDB; AAL45648, AAL45649.
XX
PT Novel human protein, and the polynucleotide that encodes it, useful for
PT inhibiting cancer cell growth.
XX
PS Claim 1; Page 14 (Disclosure); 31pp; Chinese.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human proteins capable of acting as cancer cell growth
CC inhibitors. These can be used to inhibit cancer. The present sequence is
CC a protein described in the exemplification of the invention.
XX
SQ Sequence 184 AA;
XX
Query Match 16.0%; Score 264; DB 23; Length 184;
Best Local Similarity 37.3%; Pred. No. 4.5e-19;
Matches 69; Conservative 27; Mismatches 77; Indels 12; Gaps 3;
QY 118 VEPKPKLHTGNFAFMICERRPAIQGVJLLYAIQVPSSEFLDFLTPRLRAFGCLR 197
Dh 4 ISFPRKPLRPNFTARTIAPRSPAFQFLGHLQAVPELHARDLQDFVLPRLRAQSLTC 63
QY 108 AGQYPPRAIFILPVPPLCEKI-TAHCPAAAVPALCAVLLCHPDLDRPAEAPAGEPAIQ 256
Dh 64 TGLVFEALALMANAWQLQAQVTPSPGPPPLITLAGLVCHQELDPGEAPACCEPA:QL 123
QY 257 LQAPGHHYVAPLLDAMPPLAVYALGKDFVTI:QPR-ESQLPPTPPG:T--- :KELT 310
Dh 124 LGDKSLHLPLAFLEAHVRLSWRLG-----LDKROSEARLQALQDEAGLTPPPSLKELL 178
QY 311 VREYL 315
Dh 179 IKEVL 183
XX
RESULT 6
ABG27894
ID ABG27894 standard; Protein; 580 AA.
XX
AC ABG27894;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #27885.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-0508631.
XX

```

```

PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
XX
DR N-PSDB; AAS92081.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 20; SEQ ID No 58253; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (II) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 580 AA;
XX
Query Match 15.5%; Score 255; DB 22; Length 580;
Best Local Similarity 100.0%; Pred. No. 2.1e-17;
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MASPEHSGSPGCMGPITQCTARTQGEAPATGPDLPHPGPDGHL 44
Dh 350 MASPEHSGSPGCMGPITQCTARTQGEAPATGPDLPHPGPDGHL 393
XX
RESULT 7
AAM84390
ID AAM84390 standard; Protein; 93 AA.
XX
AC AAM84390;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen SEQ ID NO:11993.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
XX

```

PR 24 FEB-2000: 2000US-0184664  
 PR 02 MAR 2000: 2000US-0186150  
 PR 16 MAR 2000: 2000US-0189874  
 PR 17 MAR 2000: 2000US-0190076  
 PR 18 APR 2000: 2000US-0194123  
 PR 19 MAY 2000: 2000US-0205515  
 PR 07 JUN 2000: 2000US-0209467  
 PR 28 JUN 2000: 2000US-0214886  
 PR 30 JUN 2000: 2000US-0215135  
 PR 07 JUL 2000: 2000US-0216647  
 PR 07 JUL 2000: 2000US-0216880  
 PR 11 JUL 2000: 2000US-0217487  
 PR 14 JUL 2000: 2000US-0217496  
 PR 26 JUL 2000: 2000US-0220963  
 PR 26 JUL 2000: 2000US-0220964  
 PR 14 AUG 2000: 2000US-0224518  
 PR 14 AUG 2000: 2000US-0225213  
 PR 14 AUG 2000: 2000US-0225214  
 PR 14 AUG 2000: 2000US-0225266  
 PR 14 AUG 2000: 2000US-0225267  
 PR 14 AUG 2000: 2000US-0225268  
 PR 14 AUG 2000: 2000US-0225270  
 PR 14 AUG 2000: 2000US-0225447  
 PR 14 AUG 2000: 2000US-0225757  
 PR 14 AUG 2000: 2000US-0225758  
 PR 14 AUG 2000: 2000US-0225759  
 PR 18 AUG 2000: 2000US-0226279  
 PR 22 AUG 2000: 2000US-0226581  
 PR 22 AUG 2000: 2000US-0226764  
 PR 22 AUG 2000: 2000US-0227182  
 PR 23 AUG 2000: 2000US-0227009  
 PR 30 AUG 2000: 2000US-0228924  
 PR 01 SEP 2000: 2000US-0229287  
 PR 01 SEP 2000: 2000US-0229343  
 PR 01 SEP 2000: 2000US-0229344  
 PR 01 SEP 2000: 2000US-0229345  
 PR 05 SEP 2000: 2000US-0229509  
 PR 05 SEP 2000: 2000US-0229713  
 PR 06 SEP 2000: 2000US-0230437  
 PR 06 SEP 2000: 2000US-0230438  
 PR 08 SEP 2000: 2000US-0231147  
 PR 08 SEP 2000: 2000US-0231243  
 PR 08 SEP 2000: 2000US-0231244  
 PR 08 SEP 2000: 2000US-0231413  
 PR 08 SEP 2000: 2000US-0231414  
 PR 08 SEP 2000: 2000US-0232080  
 PR 08 SEP 2000: 2000US-0232081  
 PR 12 SEP 2000: 2000US-0231268  
 PR 14 SEP 2000: 2000US-0232197  
 PR 14 SEP 2000: 2000US-0232198  
 PR 14 SEP 2000: 2000US-0232199  
 PR 14 SEP 2000: 2000US-0232400  
 PR 14 SEP 2000: 2000US-0232401  
 PR 14 SEP 2000: 2000US-0233063  
 PR 14 SEP 2000: 2000US-0233064  
 PR 14 SEP 2000: 2000US-0233065  
 PR 21 SEP 2000: 2000US-0234223  
 PR 21 SEP 2000: 2000US-0234224  
 PR 25 SEP 2000: 2000US-0234997  
 PR 25 SEP 2000: 2000US-0234998  
 PR 26 SEP 2000: 2000US-0235484  
 PR 27 SEP 2000: 2000US-0235834  
 PR 27 SEP 2000: 2000US-0235836  
 PR 29 SEP 2000: 2000US-0236327  
 PR 29 SEP 2000: 2000US-0236327  
 PR 29 SEP 2000: 2000US-0236367  
 PR 29 SEP 2000: 2000US-0236368  
 PR 29 SEP 2000: 2000US-0236369  
 PR 29 SEP 2000: 2000US-0236370  
 PR 02 OCT 2000: 2000US-0237032  
 PR 02 OCT 2000: 2000US-0237037  
 PR 02 OCT 2000: 2000US-0237038

PR 02 OCT 2000: 2000US-0237039  
 PR 02 OCT 2000: 2000US-0237040  
 PR 13 OCT 2000: 2000US-0238935  
 PR 13 OCT 2000: 2000US-0239937  
 PR 20 OCT 2000: 2000US-0240360  
 PR 20 OCT 2000: 2000US-0241221  
 PR 20 OCT 2000: 2000US-0241785  
 PR 20 OCT 2000: 2000US-0241786  
 PR 20 OCT 2000: 2000US-0241787  
 PR 20 OCT 2000: 2000US-0241808  
 PR 20 OCT 2000: 2000US-0241809  
 PR 20 OCT 2000: 2000US-0241825  
 PR 01 NOV 2000: 2000US-0244617  
 PR 08 NOV 2000: 2000US-0246474  
 PR 08 NOV 2000: 2000US-0246475  
 PR 08 NOV 2000: 2000US-0246476  
 PR 08 NOV 2000: 2000US-0246477  
 PR 08 NOV 2000: 2000US-0246478  
 PR 08 NOV 2000: 2000US-0246523  
 PR 08 NOV 2000: 2000US-0246524  
 PR 08 NOV 2000: 2000US-0246525  
 PR 08 NOV 2000: 2000US-0246526  
 PR 08 NOV 2000: 2000US-0246527  
 PR 08 NOV 2000: 2000US-0246528  
 PR 08 NOV 2000: 2000US-0246531  
 PR 08 NOV 2000: 2000US-0246609  
 PR 08 NOV 2000: 2000US-0246610  
 PR 08 NOV 2000: 2000US-0246611  
 PR 08 NOV 2000: 2000US-0246613  
 PR 17 NOV 2000: 2000US-0246211  
 PR 17 NOV 2000: 2000US-0246212  
 PR 17 NOV 2000: 2000US-0246213  
 PR 17 NOV 2000: 2000US-0246215  
 PR 17 NOV 2000: 2000US-0246216  
 PR 17 NOV 2000: 2000US-0246217  
 PR 17 NOV 2000: 2000US-0246218  
 PR 17 NOV 2000: 2000US-0246244  
 PR 17 NOV 2000: 2000US-0246245  
 PR 17 NOV 2000: 2000US-0246246  
 PR 17 NOV 2000: 2000US-0246247  
 PR 17 NOV 2000: 2000US-0246297  
 PR 17 NOV 2000: 2000US-0246299  
 PR 17 NOV 2000: 2000US-0246300  
 PR 01 DEC 2000: 2000US-0250160  
 PR 01 DEC 2000: 2000US-0250391  
 PR 01 DEC 2000: 2000US-0251030  
 PR 05 DEC 2000: 2000US-0251988  
 PR 05 DEC 2000: 2000US-0256719  
 PR 06 DEC 2000: 2000US-0251479  
 PR 06 DEC 2000: 2000US-0251856  
 PR 06 DEC 2000: 2000US-0251866  
 PR 08 DEC 2000: 2000US-0251869  
 PR 08 DEC 2000: 2000US-0251899  
 PR 08 DEC 2000: 2000US-0251990  
 PR 11 DEC 2000: 2000US-0254097  
 PR 05 JAN 2001: 2001US-0259678

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI, 2001-484426/52.

NIH PDB, AA57171.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
 useful for preventing, diagnosing and/or treating cancers and  
 metastasis -

BS Claim 11; SEQ ID NO 11983; 3071pp + Sequence Listing; English.  
 XX  
 CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)  
 CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic  
 CC activity, and can be used in gene therapy and vaccine production. (I)  
 CC proteins and polynucleotides may be used in the prevention, diagnosis and  
 CC treatment of diseases associated with inappropriate (I) expression. For  
 CC example, they may be used to treat disorders associated with decreased  
 CC expression by rectifying mutations or deletions in a patient's genome  
 CC that affect the activity of (I) by expressing inactive proteins or to  
 CC supplement the patient's own production of (I). Additionally, (I)  
 CC polynucleotides may be used to produce the secreted (I), by inserting  
 CC the nucleic acids into a host cell and culturing the cell to express the  
 CC protein. (I) proteins and polynucleotides may be used to prevent,  
 CC diagnose and treat immune/haematopoietic-related diseases, especially  
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
 CC to AAK67694 represent human immune/haematopoietic antigen genomic  
 CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169  
 CC represent sequences used in the exemplification of the present invention.  
 XX  
 SQ Sequence 93 AA;  
 Query Match 15.1%; Score 249; DB 22; Length 93;  
 Best Local Similarity 100.0%; Pred No. 6, 1e-18;  
 Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MASPEHPSGCMSPITQTAPTCQEPATGPDLPHPGPDGHL 43  
 Db 46 MASPEHPSGCMSPITQTARTQDEAPATGPDLPHPGPDGHL 88  
 RESULT 8  
 ABB59625  
 ID ABB59625 standard; Protein: 295 AA.  
 XX  
 AC ABB59625;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster polypeptide SEQ ID NO 5667.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 XX pharmaceutical.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 FN W0200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO US092331.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 XX  
 PR 11-JUL-2000; 2000US-0614150  
 XX  
 PA (PEKE ) PE CORP NY  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 DR WPI; 2001-656860/75.  
 XX  
 DR N-PSDB; ABL03728.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 PS Disclosure; SEQ ID NO 5667; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA  
 CC sequences (AB101840-AB116175) and the encoded proteins  
 CC (AB57737-AB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 295 AA;  
 Query Match 11.2%; Score 185.5; DB 22; Length 295;  
 Best Local Similarity 27.7%; Pred No. 1, 4e-10;  
 Matches 80; Conservative 38; Mismatches 102; Indels 69; Gaps 13;  
 QY 37 PGPDGHLDT---HSGLSNNSMTTPELQQ-YWQN-QKCPWK-----HYKLLFEIASARI 85  
 Db 16 PGPD-ELDSALPAALDIPPESDKALQGWERTASVAYKPTDGSIVLRFIDLAHI 74  
 QY 86 -----FEPKVSFVWVYQIVIQTSF-DNNKAVLEPPYSDFAKLQKALIKTPEETNVF 139  
 Db 75 MPDGDGVKIKPFVVELIVKQMGATEDTPAPKTEPPYDFFPIYIGIKPKHPAEKANY 134  
 QY 140 FPKKHITGNFAEEMICEPPALQEVYGLVYAIKVPSPPEFD---FITPEPLFEAFQC 195  
 Db 135 FPAKVMGNKFSKELLIGERGAAPFLL .....TACQFLDERRNMATPILNCFRL 184  
 QY 196 LRAGQYPRALELLRLVLPQEKLTAGCPAAV-----ALCAV---LICHRLDRPA 244  
 Db 185 LNKTYMRSRPVLLILC---RIVAACTSSPVPHNAEPMALLAUSPFETLCDIDL--- 236  
 QY 245 EAFAGSPALQPLQAPEGHYVYPIIDAWVPLAYALGKDFVTLQEPLEE 293  
 Db 237 -----LPLVILPLHTCAHLWQRGQDQKPTITRLTD 267  
 RESULT 9  
 ABG23125  
 ID ABG23125 standard; Protein: 212 AA.  
 XX  
 AC ABG23125;  
 XX  
 DT 18-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #23116.  
 XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder  
 XX  
 OS Homo sapiens.  
 XX  
 FN W0200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540317.  
 XX  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;  
 XX  
 DR WPI; 2001-639362/73.  
 XX  
 DR N-PSDB; AAS87312.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX  
 PS Claim 20, SEQ ID NO 53484, 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG10377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIP0  
 CC at ftp.wipo.int/pub/published\_pot\_sequences  
 CC  
 XX Sequence 212 AA:  
 XX  
 XX Query Match 7.6%; Score 135.5; E-Value 2.2; Length 212,  
 XX Best Local Similarity 28.2%; Pred. No 0.00016;  
 XX Matches 51; Conservative 27; Mismatches 58; Indels 45; Gaps 10;  
 XX  
 XX 24 GQEAR-ATGPDLPHPHGHLDHSHLSNSMTPELQGYWNGKCPKHKILFPLAS 82  
 XX 13 ETQWMLRPLPLPVG-----TPRAPPPGAPAPMEVY-----TPS 50  
 XX  
 XX 83 ARLEPRVSR-FVYQIIVITQTSFNNKAVLRPRYSPAKLOKALTPPEIEDEFF 141  
 XX 51 FRYESDLERGTYVFKTEVMNG---PRHFEYKPSFPHALHYKLYK ---ITPRIP 102  
 XX  
 XX 142 RKHTNCPAEMICERRALQELGLYALPVPSPF---FLPFLT---PPELRAPG 194  
 XX 103 SKHP NVWPKVLEQRPQGLTYLQAV---ILENELPKLPLDPLNPHLSPPAES 156  
 XX  
 XX 195 C 195  
 XX 157 C 157  
 XX  
 XX RESULT 10  
 XX AAU17487  
 XX ID AAU17487 standard; Protein; 173 AA  
 XX AC AAU17487;  
 XX  
 XX 07 NOV-2001 (first entry)  
 XX  
 XX Novel signal transduction pathway protein, Seq ID 1052  
 XX  
 XX Receptor, cytosolic; dermatological; immunosuppressive; tumour;  
 XX anti-inflammatory; anti HIV; antibacterial; anti-inflammatory; cancer;  
 XX immune system disorder; rheumatoid arthritis; inflammatory condition;  
 XX organ transplant rejection; infection; hepatitis C; blood disorder;  
 XX sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;  
 XX neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
 XX chromosomal abnormality; Down syndrome; ischaemia; renal disorder;  
 XX cardiovascular; respiratory; wound healing; endocrine; Addison's disease;  
 XX reproductive system; gastrointestinal; liver disorder; AIDS;  
 XX acquired immune deficiency syndrome  
 XX  
 XX Homo sapiens.  
 XX  
 XX W0200154733-A1.  
 XX  
 XX 02 AUG-2001  
 XX  
 XX 17 JAN 2001; 2001WO/050112  
 XX  
 XX

PR 31-JAN-2001; 2000US 0179065.  
 PR 04-FEB-2001; 2000US 0180629.  
 PR 24-FEB-2001; 2000US 0184664.  
 PR 02-MAR-2001; 2000US 0186350.  
 PR 16-MAR-2001; 2000US 0189874.  
 PR 17-MAR-2001; 2000US 0190076.  
 PR 18-APR-2001; 2000US 0198123.  
 PR 19-MAY-2001; 2000US 0205515.  
 PR 07-JUN-2001; 2000US 0205467.  
 PR 28-JUN-2001; 2000US 0214886.  
 PR 30-JUN-2001; 2000US 0215135.  
 PR 07-JUL-2001; 2000US 0216647.  
 PR 07-JUL-2001; 2000US 0216688.  
 PR 11-JUL-2001; 2000US 0217487.  
 PR 11-JUL-2001; 2000US 0217496.  
 PR 14-JUL-2001; 2000US 0218290.  
 PR 26-JUL-2001; 2000US 0220963.  
 PR 26-JUL-2001; 2000US 0220964.  
 PR 14-AUG-2001; 2000US 0224518.  
 PR 14-AUG-2001; 2000US 0224519.  
 PR 14-AUG-2001; 2000US 0225213.  
 PR 14-AUG-2001; 2000US 0225214.  
 PR 14-AUG-2001; 2000US 0225266.  
 PR 14-AUG-2001; 2000US 0225267.  
 PR 14-AUG-2001; 2000US 0225268.  
 PR 14-AUG-2001; 2000US 0225270.  
 PR 14-AUG-2001; 2000US 0225447.  
 PR 14-AUG-2001; 2000US 0225757.  
 PR 14-AUG-2001; 2000US 0225758.  
 PR 14-AUG-2001; 2000US 0225759.  
 PR 18-AUG-2001; 2000US 0226279.  
 PR 22-AUG-2001; 2000US 0226681.  
 PR 22-AUG-2001; 2000US 0226686.  
 PR 22-AUG-2001; 2000US 0227182.  
 PR 23-AUG-2001; 2000US 0227109.  
 PR 30-AUG-2001; 2000US 0228324.  
 PR 01-SEP-2001; 2000US 0228287.  
 PR 01-SEP-2001; 2000US 0229343.  
 PR 01-SEP-2001; 2000US 0229344.  
 PR 01-SEP-2001; 2000US 0229345.  
 PR 05-SEP-2001; 2000US 0229509.  
 PR 05-SEP-2001; 2000US 0229510.  
 PR 06-SEP-2001; 2000US 0230437.  
 PR 06-SEP-2001; 2000US 0230438.  
 PR 08-SEP-2001; 2000US 0231242.  
 PR 08-SEP-2001; 2000US 0231243.  
 PR 08-SEP-2001; 2000US 0231244.  
 PR 08-SEP-2001; 2000US 0231413.  
 PR 08-SEP-2001; 2000US 0231414.  
 PR 08-SEP-2001; 2000US 0232080.  
 PR 08-SEP-2001; 2000US 0232081.  
 PR 12-SEP-2001; 2000US 0231968.  
 PR 14-SEP-2001; 2000US 0233397.  
 PR 14-SEP-2001; 2000US 0233398.  
 PR 14-SEP-2001; 2000US 0233399.  
 PR 14-SEP-2001; 2000US 0233400.  
 PR 14-SEP-2001; 2000US 0233401.  
 PR 14-SEP-2001; 2000US 0233063.  
 PR 14-SEP-2001; 2000US 0233064.  
 PR 14-SEP-2001; 2000US 0233065.  
 PR 21-SEP-2001; 2000US 0234223.  
 PR 21-SEP-2001; 2000US 0234274.  
 PR 25-SEP-2001; 2000US 0234997.  
 PR 25-SEP-2001; 2000US 0234998.  
 PR 25-SEP-2001; 2000US 0234984.  
 PR 27-SEP-2001; 2000US 0235834.  
 PR 27-SEP-2001; 2000US 0235836.  
 PR 29-SEP-2001; 2000US 0236327.  
 PR 29-SEP-2001; 2000US 0236367.  
 PR 29-SEP-2001; 2000US 0236368.  
 PR 29-SEP-2001; 2000US 0236369.  
 PR 29-SEP-2001; 2000US 0236370.  
 PR 02-OCT-2001; 2000US 0236802.

PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239235.  
 PR 13-OCT-2000; 2000US-0239237.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241222.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 01-NOV-2000; 2000US-0241825.  
 PR 01-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249246.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251033.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 06-DEC-2000; 2000US-0256719.  
 PR 08-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC  
 XX  
 PI Rosen CA, Barash SC, Ruben SM;  
 XX WPI; 2001-465460/50.  
 DR N-PSDB; AAS27404.  
 XX  
 PT Novel polypeptides useful for diagnosing, treating, preventing and/or  
 PT prognosing disorders related to the proteins, including cancers, immune

PT disorders and neuronal disorders -  
 XX  
 XX Claim 1, SEQ ID No 1052; 880bp; English.  
 XX  
 CC The invention relates to novel isolated polypeptides (I), and  
 CC polynucleotides (II), (I), (II) and the antibody to (I) are useful for  
 CC diagnosing, preventing and treating diseases including immune system  
 CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune  
 CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ  
 CC transplant rejections and graft versus host disease, infectious diseases  
 CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and  
 CC other blood-related disorders (sickle cell anaemia), myeloproliferative  
 CC disorders, primary haematopoietic disorders, hyperproliferative  
 CC disorders (e.g. Gaucher's disease and cancer), neurodegenerative  
 CC disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal  
 CC abnormalities (Down syndrome), ischaemic injury (e.g. stroke), renal  
 CC disorders (e.g. glomerulonephritis), cardiovascular disorders  
 CC (e.g. arrhythmia), respiratory disorders, dermatological disorders, in  
 CC wound healing, epithelial cell proliferation, endocrine disorders (e.g.  
 CC Addison's disease), reproductive system disorders, gastrointestinal  
 CC disorder (inflammatory disorders), liver disorders (cirrhosis),  
 CC as stimulators of B-cell responsiveness to pathogens, activators of  
 CC T-cells, to induce higher affinity antibodies, and as a means to induce  
 CC tumour proliferation in pathologies e.g. acquired immune deficiency  
 CC syndrome (AIDS). AAU17059-AAU17683 represent novel signal transduction  
 CC pathway protein, amino acid sequences of the invention.  
 CC  
 XX

Query Match 7.2%; Score 118; DB 22; Length 173;  
 Best Local Similarity 32.3%; Pred. No. 0.00071;  
 Matches 40; Conservative 22; Mismatches 40; Indels 22; Gaps 7,

QY 80 IASAPLEERRVSK-FVYVIVITSTFNNKFAVLPYSPFAKLGKALTFPEIEDV 138  
 DB 19 IPSFRYEESDLEPSTVFTEVLWN3---PPIFEKPYSEFHALPKYK\*---ITP 70

QY 139 EPPRKHTGNAEMICERPALCEYLGLVAIFCPSPRE---FLPILT---PPELPE 191  
 DB 71 EIPSNVPR-IMWPKVLEDPKGLTYLQAV-----ILNENELPKLFLDLNVPHLPSLPK 124

QY 192 AFSC 195  
 DB 125 AESC 128

RESULT 11  
 ABB59662  
 ID ABB59662 standard; Protein, 431 AA.

AC ABB59662;

DT 26-MAR 2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 5778.

KM Drosophila: developmental biology; cell signalling; insecticide;  
 KM pharmaceutical.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001, 2001WC-US09231.

PR 23-MAR-2000; 2000US-191637P

PR 11-JUL-2000; 2000US-0614150.

PA (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PMD, Myers EW;

DR WPI: 2001-656860/75.  
 DR N-PSDB; ABL03765.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 PS Disclosure; SEQ ID NO 5778; 21pp + Sequence Listing, English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
 CC sequences (AB101840-AB11175) and the encoded proteins  
 CC (AB857731-AB872072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from Wipo  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX  
 SO Sequence 431 AA;  
 Query Match 6.8%; Score 114.5; DB 22; Length 431;  
 Best Local Similarity 25.9%; Pred. No. 0.0063;  
 Matches 44; Conservative 20; Mismatches 63; Indels 43; Gaps 6,  
 QY 23 TGGCAPATGTPGTPGTPGTHLTHSLSSNSMTTELQYVWQKCRKMKVKKLEFIAS 82  
 Db 41 TGGACADLEIQ-----EDSGSYIDYDKSLP-----ISLPD 74  
 QY 83 ARLEERVSPFVVVQIVIGTSPFNKNKAVLERRYSDFALQKALKTPEEIEVEFPR 142  
 Db 75 YGIVNPRMEREYIVFNHMA-----GPGICSPRPEFANLHSL-----PREFSGNPRY 123  
 QY 143 KHLTGNF---AEEMLTEPPPAIGFYGLVYAFVCPSPSPFFLPIPTPE 184  
 Db 124 --LPGKWPQLSEQQILNPPRGLQYLEKVAVPVIAESNAVQDFLPTDE 171  
 RESULT 12  
 AAB45187  
 ID AAB45187 standard; Protein; 168 AA.  
 AC  
 XX AAB45187;  
 XX  
 DT 12 FEB-2001 (first entry)  
 XX  
 DE Human secreted protein sequence encoded by gene 15 SEQ ID NO:128.  
 KW Human; secreted protein; immunosuppressive; antiarthritic; antirheumatic;  
 KW antiproliferative; cyostatic; cardiant; vasotrophic; cerebroprotective;  
 KW neurotropic; neuroprotective; antibacterial; vituicide; fungicide; cancer;  
 KW ophthalmological; autoimmune disease; hyperproliferative disorder;  
 KW cardiovascular disorder; cerebrovascular disorder; wound healing;  
 KW nervous system disorder; aging; chemotaxis.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200058467-A1.  
 PD 05-OCT-2000.  
 XX  
 DE 22 MAR 2000; JMOUW-18507505  
 XX  
 PR 26 MAR 1999; 9408-0126502  
 PR 17 DEC 1999; 9908-012410.  
 XX  
 PA (HIMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Ruben SM, Komatsoulis G;  
 XX  
 DR WPI: 2000-611712/58.

DR N-PSDB; AAC80545.  
 XX  
 PT Nucleic acid molecules encoding human secreted proteins, used in  
 PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and  
 PT Parkinson's diseases and cancers -  
 XX  
 PS Disclosure; Page 28; 440pp; English.  
 XX  
 CC Polynucleotide sequences AAC80531-C80580 represent cDNA encoding human  
 CC secreted proteins AAB45120-B45169. Sequences AAB45170-B45225 represent  
 CC alternative polypeptides encoded by the genes, and amino acid sequences  
 CC to which they are homologous. The genes and proteins have activities  
 CC dependent on the tissues and cells in which they are expressed. Examples  
 CC of their activities include immunosuppressive; antirheumatic;  
 CC antithematic; antiproliferative; cyostatic; cardiant; vasotrophic;  
 CC cerebroprotective; neurotropic; neuroprotective; antibacterial; vituicide;  
 CC fungicide; and ophthalmological. The secreted proteins, polynucleotides,  
 CC antagonists and agonists may be useful in treating, preventing and/or  
 CC diagnosing diseases and disorders such as autoimmune diseases  
 CC e.g. rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms  
 CC of the breast or liver, cardiovascular disorders e.g. cardiac arrest,  
 CC cerebrovascular disorders e.g. cerebral ischaemia, anglogenesis, nervous  
 CC system disorders e.g. Alzheimer's disease, infections caused by  
 CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection.  
 CC The polypeptides can also be used to aid wound healing and epithelial  
 CC cell proliferation, to prevent skin aging due to sunburn, to maintain  
 CC organs before transplantation, for supporting cell culture of primary  
 CC tissues, to regenerate tissues and in chemotaxis. The polypeptides can  
 CC also be used as a food additive or preservative to increase or decrease  
 CC storage capabilities. AAC80522-cranin and AAB45119 represent sequences  
 CC used in the isolation and characterisation of the genes and proteins of  
 CC the invention.  
 XX  
 SO Sequence 168 AA;  
 Query Match 6.8%; Score 112.5; DB 21; Length 168;  
 Best Local Similarity 29.5%; Pred. No. 0.0025;  
 Matches 35; Conservative 22; Mismatches 37; Indels 23; Gaps 5;  
 QY 70 RMKVKKLLEFIASAPTEPKVSKFVVVQIVIGTSPFNKNKAVLEPRYSPFAKQKALK 129  
 Db 40 RYKRV-----EOMGEKPFVVVNVMA-----GQULSKRPEFALHQNVL-- 78  
 QY 130 TPFETEDVEFPFKHLTG---NFAEMICEPPPAIGFYGLVYAFVCPSPSPFFLPIPT 185  
 Db 79 --KREFANFTPP--LPGKWPQLSEQQILNPPRGLQYLEKVAVPVIAESNAVQDFLPTDE 134  
 QY 186 RPE 188  
 Db 135 ESD 137  
 RESULT 13  
 ABB08158  
 ID ABB08158 standard; Protein; 541 AA.  
 AC  
 XX ABB08158;  
 XX  
 DT 10-SEP-2002 (first entry)  
 XX  
 DE Human cytoskeleton-associated protein (CSAP)-2 (ID: 959690CTD1).  
 KW Cytoskeleton-associated protein; CSAP; antiarteriosclerotic; human;  
 KW hepatotropic; antiinflammatory; antipsoriasis; cyostatic; vituicide;  
 KW anticonvulsant; cerebroprotective; neurotropic; neuroprotective;  
 KW antiparkinsonian; transgenic; gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200242330-A2.  
 PD 30-MAY-2002.  
 XX



26-OCT-2001; 2001WO-US0983.  
 XX 27-OCT-2000; 2000US-244022P.  
 PR 08-NOV-2000; 2000US-247370P.  
 PR 07-DEC-2000; 2000US-251831P.  
 XX  
 PA (INCYTE GENOMICS INC.  
 XX  
 PI Baughn MR, Yao MG, Walla NK, Gietzen KJ, Thangavelu K, Lu Y;  
 PI Ding L, Yue H, Tang YT, Lal PG, Batra S, Lu DM, Sanjanwala MM;  
 PI Arvizu C, Ramkumar J, Griffin JA, Gururajan R, Azimzai Y, Xu Y;  
 PI Burford N;  
 XX  
 DR WPI: 2002-500279/53.  
 DR N-PSDE: ABLK0868  
 XX  
 PT Human cytoskeleton-associated protein, useful in diagnosis, prevention  
 PT or treatment of cell proliferative disorders, viral infections and  
 PT neurological disorders  
 XX  
 PS Claim 1; Page 118-119; 163pp; English.  
 XX  
 CC The invention relates to human cytoskeleton-associated proteins (CSAP)  
 CC and encoding polynucleotides. The CSAP polypeptides can be expressed by  
 CC standard recombinant methodology. The CSAP polynucleotides, polypeptides  
 CC and modulators are useful for diagnosing, treating or preventing  
 CC disorders associated with aberrant expression of CSAP, where the  
 CC disorders are selected from a cell proliferative disorder such as  
 CC arteriosclerosis, cirrhosis, hepatitis, psoriasis, cancer, a viral  
 CC infection such as those caused by adenoviruses, herpesviruses, and  
 CC retroviruses, and a neurological disorder such as epilepsy, stroke,  
 CC Alzheimer's disease, Huntington's disease, dementia, and Parkinson's  
 CC disease. The polypeptides are also useful in a number of drug screening  
 CC techniques, and to analyse the proteome of a tissue or cell type. The  
 CC present sequence represents a human CSAP-2 polypeptide.  
 XX  
 SO Sequence 541 AA;  
 Query Match 6.8%; Score 112.5; DB 23; Length 541;  
 Best Local Similarity 28.5%; Pred No 0.014;  
 Matches 35, Conservative 23, Mismatches 37, Indels 29, Gaps 5,  
 QY 70 PKMKYKLEIFIASAPLEPKVSPVYVQIVITGSPNNFAVLEPPSPDPAVLQVLLX 129  
 DB 166 PKYHV-----FNDKDFVYVYVMA-----DPGLCKRPREPAILQNL 207  
 QY 130 TPEEIEIDVPPPKHJTG---NFAEEMICEPPALQEVLSGLVAIPVPPSPPEFLPFT 186  
 DB 209 --YPPFANFTFP--LPGWPPSLSECGIDAPRPGIEELKLVKSPVYGESDINGCEPLS 263  
 QY 166 RPE 188  
 DB 264 ESD 266  
 RESULT 14  
 AAU17481  
 ID AAU17481 standard; Protein; 141 AA.  
 XX  
 AC AAU17481;  
 XX  
 DT 07-NOV-2001 (first entry)  
 XX  
 DE Novel signal transduction pathway protein, Seq ID 1046.  
 XX  
 KM Neutropenic, cytostatic, dermatological, immunosuppressive, tumour;  
 KM anti-inflammatory; anti-HIV; anti-bacterial; anti-inflammatory; cancer;  
 KM immune system disorder; rheumatoid arthritis; inflammatory condition;  
 KM organ transplant rejection; infection; hepatitis C; blood disorder;  
 KM sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;  
 KM neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
 KM chromosomal abnormality; Down syndrome; ischaemia; renal disorder;  
 KM cardiovascular; respiratory; wound healing; endocrine; Addison's disease;

reproductive system, gastrointestinal, liver disorder, AIDS,  
 acquired immune deficiency syndrome.  
 XX Homo sapiens.  
 XX MO200154733-A1.  
 XX  
 PD 02-AUG-2001.  
 XX  
 PF 17-JAN-2001; 2001WO-US01312.  
 XX  
 PR 31-JAN-2000; 2000US-0179065.  
 PR 04-FEB-2000; 2000US-0180628.  
 PR 24-FEB-2000; 2000US-0184664.  
 PR 02-MAR-2000; 2000US-0186350.  
 PR 16-MAR-2000; 2000US-0189874.  
 PR 17-MAR-2000; 2000US-0190076.  
 PR 18-APR-2000; 2000US-0198123.  
 PR 19-MAY-2000; 2000US-0205515.  
 PR 07-JUN-2000; 2000US-0209467.  
 PR 28-JUN-2000; 2000US-0214886.  
 PR 30-JUN-2000; 2000US-0215135.  
 PR 07-JUL-2000; 2000US-0216647.  
 PR 07-JUL-2000; 2000US-0216880.  
 PR 11-JUL-2000; 2000US-0217487.  
 PR 11-JUL-2000; 2000US-0217496.  
 PR 14-JUL-2000; 2000US-0218290.  
 PR 26-JUL-2000; 2000US-0220963.  
 PR 26-JUL-2000; 2000US-0220964.  
 PR 14-AUG-2000; 2000US-0224518.  
 PR 14-AUG-2000; 2000US-0224519.  
 PR 14-AUG-2000; 2000US-0225213.  
 PR 14-AUG-2000; 2000US-0225214.  
 PR 14-AUG-2000; 2000US-0225266.  
 PR 14-AUG-2000; 2000US-0225267.  
 PR 14-AUG-2000; 2000US-0225268.  
 PR 14-AUG-2000; 2000US-0225270.  
 PR 14-AUG-2000; 2000US-0225447.  
 PR 14-AUG-2000; 2000US-0225757.  
 PR 14-AUG-2000; 2000US-0225758.  
 PR 14-AUG-2000; 2000US-0225759.  
 PR 18-AUG-2000; 2000US-0226279.  
 PR 22-AUG-2000; 2000US-0226681.  
 PR 22-AUG-2000; 2000US-0226688.  
 PR 22-AUG-2000; 2000US-0227182.  
 PR 22-AUG-2000; 2000US-0227182.  
 PR 23-AUG-2000; 2000US-0227009.  
 PR 30-AUG-2000; 2000US-0228824.  
 PR 01-SEP-2000; 2000US-0229287.  
 PR 01-SEP-2000; 2000US-0229343.  
 PR 01-SEP-2000; 2000US-0229344.  
 PR 01-SEP-2000; 2000US-0229345.  
 PR 05-SEP-2000; 2000US-0229509.  
 PR 05-SEP-2000; 2000US-0229513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 06-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 08-SEP-2000; 2000US-0232081.  
 PR 12-SEP-2000; 2000US-0231968.  
 PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 14-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 21-SEP-2000; 2000US-0234274.

PR 25-SEP-2000: 2000US-0234997  
 PR 25-SEP-2000: 2000US-0234998.  
 PR 26-SEP-2000: 2000US-0234984.  
 PR 27-SEP-2000: 2000US-0235834.  
 PR 27-SEP-2000: 2000US-0235836.  
 PR 29-SEP-2000: 2000US-0236327.  
 PR 29-SEP-2000: 2000US-0236367.  
 PR 29-SEP-2000: 2000US-0236468.  
 PR 29-SEP-2000: 2000US-0236469.  
 PR 29-SEP-2000: 2000US-0236370.  
 PR 02-OCT-2000: 2000US-0236802.  
 PR 02-OCT-2000: 2000US-0237037.  
 PR 02-OCT-2000: 2000US-0237038.  
 PR 02-OCT-2000: 2000US-0237040.  
 PR 02-OCT-2000: 2000US-0237049.  
 PR 13-OCT-2000: 2000US-0239935.  
 PR 13-OCT-2000: 2000US-0239937.  
 PR 20-OCT-2000: 2000US-0240960.  
 PR 20-OCT-2000: 2000US-0241221.  
 PR 20-OCT-2000: 2000US-0241785.  
 PR 20-OCT-2000: 2000US-0241786.  
 PR 20-OCT-2000: 2000US-0241787.  
 PR 28-OCT-2000: 2000US-0241808.  
 PR 28-OCT-2000: 2000US-0241809.  
 PR 20-OCT-2000: 2000US-0241826.  
 PR 01-NOV-2000: 2000US-0244617.  
 PR 08-NOV-2000: 2000US-0244674.  
 PR 08-NOV-2000: 2000US-0244675.  
 PR 08-NOV-2000: 2000US-0244676.  
 PR 08-NOV-2000: 2000US-0244677.  
 PR 08-NOV-2000: 2000US-0244678.  
 PR 08-NOV-2000: 2000US-0244679.  
 PR 08-NOV-2000: 2000US-0246524.  
 PR 08-NOV-2000: 2000US-0246525.  
 PR 08-NOV-2000: 2000US-0246526.  
 PR 08-NOV-2000: 2000US-0246527.  
 PR 08-NOV-2000: 2000US-0246528.  
 PR 08-NOV-2000: 2000US-0246532.  
 PR 08-NOV-2000: 2000US-0246609.  
 PR 08-NOV-2000: 2000US-0246610.  
 PR 08-NOV-2000: 2000US-0246611.  
 PR 08-NOV-2000: 2000US-0246613.  
 PR 17-NOV-2000: 2000US-0249207.  
 PR 17-NOV-2000: 2000US-0249208.  
 PR 17-NOV-2000: 2000US-0249209.  
 PR 17-NOV-2000: 2000US-0249210.  
 PR 17-NOV-2000: 2000US-0249211.  
 PR 17-NOV-2000: 2000US-0249212.  
 PR 17-NOV-2000: 2000US-0249213.  
 PR 17-NOV-2000: 2000US-0249214.  
 PR 17-NOV-2000: 2000US-0249215.  
 PR 17-NOV-2000: 2000US-0249216.  
 PR 17-NOV-2000: 2000US-0249217.  
 PR 17-NOV-2000: 2000US-0249218.  
 PR 17-NOV-2000: 2000US-0249244.  
 PR 17-NOV-2000: 2000US-0249245.  
 PR 17-NOV-2000: 2000US-0249246.  
 PR 17-NOV-2000: 2000US-0249297.  
 PR 17-NOV-2000: 2000US-0249299.  
 PR 17-NOV-2000: 2000US-0249300.  
 PR 01-DEC-2000: 2000US-0250160.  
 PR 01-DEC-2000: 2000US-0250391.  
 PR 05-DEC-2000: 2000US-0251030.  
 PR 05-DEC-2000: 2000US-0251988.  
 PR 05-DEC-2000: 2000US-0256719.  
 PR 06-DEC-2000: 2000US-0251479.  
 PR 08-DEC-2000: 2000US-0251866.  
 PR 08-DEC-2000: 2000US-0251869.  
 PR 08-DEC-2000: 2000US-0251989.  
 PR 08-DEC-2000: 2000US-0251990.  
 PR 11-DEC-2000: 2000US-0254097.

PR 05-JAN-2001, 2001US-0259678.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Barash SC, Ruben SM;  
 XX  
 DR WPI: 2001-465460/50.  
 DR N PSDB, MMS27398.  
 XX  
 PT Novel polypeptides useful for diagnosing, treating, preventing and/or  
 PT prognosing disorders related to the proteins, including cancers, immune  
 PT disorders and neuronal disorders  
 XX  
 PS Claim 1, SEQ ID No 1046, 880pp, English.  
 XX  
 CC The invention relates to novel isolated polypeptides (I), and  
 CC polynucleotides (II), (I), (II) and the antibody to (I) are useful for  
 CC diagnosing, preventing and treating diseases including immune system  
 CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune  
 CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ  
 CC transplant rejections, and graft versus host disease, infectious diseases  
 CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and  
 CC other blood-related disorders (ickle cell anaemia), myeloproliferative  
 CC disorders (e.g. Gaucher's disease and cancer), hyperproliferative  
 CC disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal  
 CC abnormalities (Down syndrome), ischemic injury (e.g. stroke), renal  
 CC disorders (e.g. glomerulonephritis), cardiovascular disorders  
 CC (e.g. arrhythmia), respiratory disorders, dermatological disorders, in  
 CC wound healing, epithelial cell proliferation, endocrine disorders (e.g.  
 CC Addison's disease), reproductive system disorders, gastrointestinal  
 CC disorder (inflammatory disorders), liver disorders (cirrhosis),  
 CC as stimulators of B-cell responsiveness to pathogens, activators of  
 CC T-cells, to induce higher affinity antibodies, and as a means to induce  
 CC tumour proliferation in pathologies e.g. acquired immune deficiency  
 CC syndrome (AIDS). AA017653-AA017683 represent novel signal transduction  
 CC pathway protein, amino acid sequences of the invention.  
 XX

Query Match 6.64, Score 108.5, DB 22, Length 141;  
 Best Local Similarity 25.94, Pred. NO. 0.0051, Indels 33, Gaps 6;  
 Matches 41, Conservative 21, Mismatches 63, Indels 33, Gaps 6;  
 QY 14 GPTTCTAFTTQEA-----PATSP-ELPHHTFQHLPTHTSSTSSSMTTPELQVWVNW 67  
 Db 8 QPSQSGSPVPDEAGARSPETVTPATQSQTSDDSL-----SDEPISNPAIINW... 57  
 QY 68 KCRKHKVCLFEIASARLEENKVKFVYQIVLTSSFNHKAVALERRISDAKLNAL 127  
 Db 58 -----IPSVFLRCRAKAAFFVYQYI-----RINDEWNIYPTPEPSLHKL 101  
 QY 128 LKTFREIEDVEFPRKHLTGFAEMICERRALQELV 165  
 Db 102 QNKY-PGVRAVNFPRKVAISNPKAFVEPPPKVQNL 138

RESULT 15  
 ABG22895  
 ID ABG22895 standard; Protein; 166 AA.  
 AC ABG22895,  
 DT 18-FEB-2002 (first entry)  
 XX  
 XX Novel human diagnostic protein #22896.  
 DE  
 XX Human, chromosome mapping; gene mapping, gene therapy, forensic;  
 KW food supplement, medical imaging; diagnostic; genetic disorder;  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.





GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: April 20, 2003, 14:07:11 ; Search time 40 Seconds  
(without alignments)  
597 350 Million cell updates/sec

Title: US-09-816-697A-2  
Perfect score: 1650

Sequence: 1 MASEFHGSGCWPITQCTAFTQCEAFATGPDLPFGDGLDTHSGUSSNMTREL 316

Scoring table: BIOSUM62  
Gap 10.0, Gapext 0.5

Searched: 288829 seqs, 75613885 residues

Total number of hits satisfying chosen parameters: 288829

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

## Database:

Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubppa/US08\_NEW\_PUB pep.\*
- 2: /cgn2\_6/ptodata/1/pubppa/PCIT\_NEW\_PUB pep.\*
- 3: /cgn2\_6/ptodata/1/pubppa/US06\_NEW\_PUB pep.\*
- 4: /cgn2\_6/ptodata/1/pubppa/US07\_NEW\_PUB pep.\*
- 5: /cgn2\_6/ptodata/1/pubppa/US07\_PUBCOMB pep.\*
- 6: /cgn2\_6/ptodata/1/pubppa/US07\_PUBCOMB pep.\*
- 7: /cgn2\_6/ptodata/1/pubppa/US08\_PUBCOMB pep.\*
- 8: /cgn2\_6/ptodata/1/pubppa/US08\_PUBCOMB pep.\*
- 9: /cgn2\_6/ptodata/1/pubppa/US09\_NEW\_PUB pep.\*
- 10: /cgn2\_6/ptodata/1/pubppa/US09\_PUBCOMB pep.\*
- 11: /cgn2\_6/ptodata/1/pubppa/US10\_NEW\_PUB pep.\*
- 12: /cgn2\_6/ptodata/1/pubppa/US10\_PUBCOMB pep.\*
- 13: /cgn2\_6/ptodata/1/pubppa/US60\_NEW\_PUB pep.\*
- 14: /cgn2\_6/ptodata/1/pubppa/US60\_PUBCOMB pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1650	100.0	316	10	US-09-816-697-2
2	118	7.2	173	9	US-09-764-868-1052
3	108.5	6.6	141	9	US-09-764-868-1046
4	99	6.0	141	9	US-09-764-868-1035
5	98.5	6.0	66	9	US-09-764-868-1182
6	96.5	6.0	165	9	US-09-764-868-1003
7	98	5.4	278	9	US-09-764-868-1051
8	95.5	5.8	2930	9	US-09-957-837A-2
9	93.5	5.7	276	9	US-09-764-868-811
10	93.5	5.7	496	10	US-09-784-249-2
11	92.5	5.6	1066	10	US-09-810-150-2
12	89	5.4	1162	10	US-09-815-242-11828
13	88	5.3	1045	10	US-09-815-242-10617
14	87.5	5.3	1993	9	US-10-098-979-2
15	85.5	5.2	621	12	US-10-042-417-28
16	85	5.2	241	10	US-09-925-302-619
17	83	5.0	343	9	US-09-938-901-6
18	83	5.0	923	9	US-09-820-843A-112
19	83	5.0	1273	9	US-09-739-330-11

20	83	5.0	1326	9	US-09-231-417-15	Sequence 15, Appl
21	83	5.0	1332	9	US-09-789-390-7	Sequence 7, Appl
22	82.5	5.0	468	10	US-09-925-300-1620	Sequence 1620, Ap
23	82.5	5.0	1244	9	US-09-789-390-13	Sequence 13, Appl
24	82.5	5.0	1303	9	US-09-789-390-9	Sequence 9, Appl
25	82	5.0	382	10	US-09-881-752A-224	Sequence 224, Appl
26	81	4.9	1945	9	US-09-927-597-2	Sequence 2, Appl
27	81	4.9	1979	9	US-09-927-597-4	Sequence 4, Appl
28	80.5	4.9	366	10	US-09-887-569A-2	Sequence 2, Appl
29	80.5	4.9	725	9	US-10-108-605-311	Sequence 321, Ap
30	80	4.8	1263	10	US-09-971-309-6	Sequence 6, Appl
31	79.5	4.8	597	9	US-09-778-319-2	Sequence 2, Appl
32	78.5	4.8	1295	9	US-09-789-390-10	Sequence 30, Appl
33	78.5	4.8	1295	9	US-09-789-390-32	Sequence 32, Appl
34	78.5	4.8	1295	9	US-09-789-390-34	Sequence 34, Appl
35	78.5	4.8	1295	9	US-09-789-390-37	Sequence 37, Appl
36	78.5	4.8	1295	9	US-09-789-390-39	Sequence 39, Appl
37	78.5	4.8	1303	9	US-09-789-390-35	Sequence 35, Appl
38	78.5	4.8	1303	9	US-09-789-390-38	Sequence 38, Appl
39	77.5	4.7	816	9	US-10-080-114A-7	Sequence 7, Appl
40	77.5	4.7	1249	9	US-09-964-899-33	Sequence 33, Appl
41	77.5	4.7	2743	9	US-10-037-182-36	Sequence 36, Appl
42	77.5	4.7	3695	9	US-10-037-182-2	Sequence 2, Appl
43	77	4.7	327	10	US-09-764-853-615	Sequence 615, App
44	76	4.6	454	9	US-09-814-604-3	Sequence 3, Appl
45	76	4.6	454	10	US-09-737-727-2	Sequence 2, Appl

## ALIGNMENTS

RESULT 1  
US-09-816-697-2  
; Sequence 2, Application US/0981697  
; Patent No. US20020037840A1  
; GENERAL INFORMATION:  
; APPLICANT: Lorenz, M., et al.  
; TITLE OF INVENTION: A NOVEL F-SEFECTIN GUT-PROTEIN LIQUANT (FSG-1)  
; FILE REFERENCE: GPN-5380  
; CURRENT APPLICATION NUMBER: US/09/816,697  
; CURRENT FILING DATE: 2001 03-23  
; PRIOR APPLICATION NUMBER: 60/192,104  
; PRIOR FILING DATE: 2000-03-24  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: Patentn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 316  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-816-697-2

Query Match 100.0%; Score 1650; DB 10; Length 316;  
Best Local Similarity 100.0%; Pred. No. 5.8e-149;  
Matches 316; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MASEFHGSGCWPITQCTAFTQCEAFATGPDLPFGDGLDTHSGUSSNMTREL	60
DB	1	MASEFHGSGCWPITQCTAFTQCEAFATGPDLPFGDGLDTHSGUSSNMTREL	60
QY	61	QQWQWQKQKWKVIVLFFIASAPIERKVFVYQIVITGSGFENNAVLEPPYSD	120
DB	61	QQWQWQKQKWKVIVLFFIASAPIERKVFVYQIVITGSGFENNAVLEPPYSD	120
QY	121	AKQKALRTFFFEETDVEFFRGLTQNAFEMTTEFFALCEYVLTLYATVYFSEFF	180
DB	121	AKQKALRTFFFEETDVEFFRGLTQNAFEMTTEFFALCEYVLTLYATVYFSEFF	180
QY	181	LDTLTPPELSEARVCFAPQYPALELLPVLPVLEXTLTAHPLAAVPAVLCVILLCHP	240
DB	181	LDTLTPPELSEARVCFAPQYPALELLPVLPVLEXTLTAHPLAAVPAVLCVILLCHP	240
QY	241	TPFAEAFAAGSEPAWICAPEDHFFVYAPILLVMPVLAVALHFFVTLAERLESCLEPPT	300

```

Db 241 DRPAENFAAGERLALQRLQAREGIRYAPILDMVPLAVALGPFVTLQEPLEESQLEPPT 300
Qy 301 PPGITLKEITVPEYH 316
Db 301 PPGITLKEITVPEYH 316

```

```

RESULT 2
US-09-764-868-1052

```

```

; Sequence 1052, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; PRIOR APPLICATION DATE: 2001-01-17
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1052
; LENGTH: 173
; TYPE: PRT
; ORGANISM: Homo sapiens
US 09 764-868-1052

```

```

Query Match
Best Local Similarity 32.3%; Score 118; DB 9; Length 173;

```

```

Matches 40; Conservative 22; Mismatches 40; Indels 22; Gaps 7,

```

```

Qy 80 TASAIEEPKVSF-PVYQIVITQTSFNNKAVLEPRYSDFAKLQKALITFEETEDV 138
Db 19 IPSFVEESDLERGVTVPKIEVLMNG---PGHVEKPYSEPHALKKLC---IKTP 70
Qy 139 EPRKHITGNFAEMICERRALQEVLLVLRCSRRE---FLDELT---RPELRE 191
Db 71 EISKSHVR-NWVPRVLEPRQGLLEYLQAV-----ILENEELPKFLDPLNVRLHLSLPK 124

```

```

Qy 192 AFQV 195
Db 125 AEGC 128

```

```

RESULT 3
US-09-764-868-1046
; Sequence 1046, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; PRIOR APPLICATION DATE: 2001-01-17
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1046
; LENGTH: 141
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: SITE
; LOCATION: (69)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US 09 764-868-1046

```

```

Query Match
Best Local Similarity 25.9%; Score 108.5; DB 9; Length 141;
Matches 41; Conservative 21; Mismatches 63; Indels 33; Gaps 6;

```

```

Qy 14 GPITQCTARTQQA-----PATGP-DLPHGPGCHLDTHSGSSMSMTTELQYQWQNO 67
Db 14 GPITQCTARTQQA-----PATGP-DLPHGPGCHLDTHSGSSMSMTTELQYQWQNO 67

```

```

Db 8 GPGSGSPVPEPEAGASRPGTVPCNLQCTSEPNQL-----SPFISNPALINW--- 57
Qy 68 KCRWPHVYLLFEIASAPIEERKVSFVYQIVITQTSFNNKAVLEPRYSDFAKLQKAL 127
Db 58 -----IPSVFLRGKAAKAFHYQYVI-----RIKDEMNIVRYRTERRSLHHKL 101
Qy 128 LKTFREIEDVEFPKHLTGNFAEMICERRALQEVLL 165
Db 102 QNKY-PQVAVVFPKKATVGNDAFVEPRKQLQNVL 138

```

```

RESULT 4
US-09-764-868-1035

```

```

; Sequence 1035, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; PRIOR APPLICATION DATE: 2001-01-17
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1035
; LENGTH: 141
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: SITE
; LOCATION: (124)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L amino acids
US-09-764-868-1035

```

```

Query Match
Best Local Similarity 33.7%; Score 99; DB 9; Length 141;
Matches 30; Conservative 12; Mismatches 39; Indels 14; Gaps 4;

```

```

Qy 82 SARIEPKVSFVYQIVITQTSFNNKAVLEPRYSDFAKLQKALITFEETEDV 139
Db 29 TAAIESSGSLQSHTEYIIRVGRGISVENSQIVKRYSDFDLNSL-----QIAGSLP 82
Qy 140 FPRKHLTGNFAEMICERRALQEVLL 168
Db 83 LPPKHLGIMDEPFIAPRQKGLQNVYLVNI 111

```

```

RESULT 5
US-09-764-868-1182

```

```

; Sequence 1182, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; PRIOR APPLICATION DATE: 2001-01-17
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1182
; LENGTH: 66
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-868-1182

```

```

Query Match
Best Local Similarity 36.5%; Score 98.5; DB 9; Length 66;
Matches 23; Conservative 11; Mismatches 24; Indels 5; Gaps 1;

```

```

Qy 95 VVOIVITQTSFNNKAVLEPRYSDFAKLQKALITFEETEDVEFPKHLTGNFAEM 154
Db 7 LVTIVIRGPPDQCPAQISRYSDFFRLHBNIQPQPGMAAISPPKRL-----REIL 61

```

QY 155 CER 157  
Db 62 LQR 64

## RESULT 6

US-09-764-868-1003  
; Sequence 1003, Application US/09764868  
; Patent No. US20020168711A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PT232  
; CURRENT APPLICATION NUMBER: US/09/764,868  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - refer to PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 1510  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1003  
; LENGTH: 165  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-764-868-1003

Query Match 6.0%; Score 98.5; DB 9; Length 165;  
Best Local Similarity 42.1%; Pred. No. 0.09; Mismatches 24; Conservative 10; Indels 1; Gaps 1;

QY 115 RPYSDPAKLOKALIKTFPEIEDVEFPFPHLTGTFNFAEMICERRALOEYLGLLVAI 171  
Db 67 RPYSPFPEMHKT-LKLVKVALALEFPFKLFGKDERVIAERSHLEKYLDPFSV 122

## RESULT 7

US-09-764-868-1051  
; Sequence 1051, Application US/09764868  
; Patent No. US20020168711A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PT232  
; CURRENT APPLICATION NUMBER: US/09/764,868  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - refer to PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 1510  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1051  
; LENGTH: 278  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (1)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (6)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (17)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (250)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-764-868-1051

## Query Match

5.9%; Score 98; DB 9; Length 278;  
Best Local Similarity 23.3%; Pred. No. 0.21; Mismatches 44; Conservative 27; Indels 56; Gaps 7;

QY 7 PGSGGCMGPI-----TQCTARTQCEAPATGPDLPHRG-----PDG 41  
Db 13 PGSTXADPPLPAAAVGAABEADEADPPASDLPTPOAIEQATVQVPAASRMQMPQG 72

QY 42 H-LDTHSGLSNSSMTTRELQOYWQOKCRWGVKLLFEIASRIEERKVSFVYQII 99  
Db 73 NPLLSHT-LQELLARDVQVLEIPEKKGLFLKIVE--YEVSSQRF----- 115  
QY 100 VIQTSFDDNNKAVLERPYSDPAKLOKALIKTFPEIEDVEFPFPHLTGTFNFAEMICERR 159  
Db 116 -----KSSVYRPYNDVVFQEMLLHKFPYPMVPAI.PPKMILAD--PEFTIAPR 163  
QY 160 ALQEYLGILL 168  
Db 164 ALKRFVLTIV 172

## RESULT 8

US-09-957-837A-2  
; Sequence 2, Application US/09957837A  
; Publication No. US20030023055A1  
; GENERAL INFORMATION:  
; APPLICANT: LOUGHNEY ET AL  
; TITLE OF INVENTION: ATR-2 CELL CYCLE CHECKPOINT  
; FILE REFERENCE: 27866/37760  
; CURRENT APPLICATION NUMBER: US/09/957,837A  
; CURRENT FILING DATE: 2001-09-21  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 2930  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-957-837A-2

Query Match 5.8%; Score 95.5; DB 9; Length 2930;  
Best Local Similarity 22.7%; Pred. No. 9.8; Mismatches 70; Conservative 48; Indels 112; Gaps 15;

QY 48 GLSSNSMTTRELQOYWQOKCRWGVKLLFEIASRIEERKVSF-----VVQIIYIQT 103  
Db 1711 GQOAEKQSKREMER-----EITPSLSSVPAELIKVMFYNRDEMLVLEPKLD 1758  
QY 104 GSEFNNKAVLERPYSDPAKLOKALIK--TFPEIEDVEFPFPHLTGTFNFAER-MTGPFP 160  
Db 1759 GSLDEVLS-LQQLTDVTELQGLKEIEFLEGAEGVDHPSHILOHRYSEHTQLOQORA 1817  
QY 161 LQYVGLALAINCVRPSPEFLDFTLPPEI.PPAFGCLPAAGYPPALDELLEV.PIQFPI 220  
Db 1818 VQEAICV-----KLNEFEOWIT--HYQAAFNNTLEATQ---LASILOEISTQMDLG- 1862  
QY 221 HCPAAVPAALC-----AVLLCHPDLDPAPAFAGEPALOPLOAPBEGHPYVAPLID 271  
Db 1863 --PSTVVPATAPLQNNAGCAHLISQCEQ--LBEGVGALLQORRSVLRGCCLEOULHY----- 1913  
QY 272 AMVRLAYVALG-----KDFVTLQERL-----ESQLRPPTP---RGIT 305  
Db 1914 ATVALQYPAIFQNIREEQMTWMEELLICNTIVERCQELYRKVMQYAPQDPPTVCCFIT 1973  
QY 306 LKELTYREY 314  
Db 1974 ATEMTLQRY 1982

## RESULT 9

US-09-764-868-811  
; Sequence 811, Application US/09764868  
; Patent No. US20020168711A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PT232  
; CURRENT APPLICATION NUMBER: US/09/764,868  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - refer to PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 1510

Query Match 5.9%; Score 98; DB 9; Length 278;  
Best Local Similarity 23.3%; Pred. No. 0.21; Mismatches 44; Conservative 27; Indels 56; Gaps 7;

```

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 811
; LENGTH: 276
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (223)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-868-811

```

```

Query Match          5.7%; Score 93.5; DB 9; Length 276;
Best Local Similarity 31.0%; Pred. No. 0.55;
Matches 27; Conservative 17; Mismatches 36; Indels 7; Gaps 3;

```

```

QY 80 IASAPTEFPVSVVYVQIIVITGSGFDNNKAVLEPPYSDFPKLCKALLKTFREIEDEVE 139
DB 18 IIPSDHEHREKKRPVYKIV-----SVGRSEMFVPRVAEFDKLYNTLKQF PAMALK 71
QY 140 FPRKHITG-NFAEEMICEPPALQEVLT 165
DB 72 IIPKPIFGNFDPPFIKQPPAGLNFTI 98

```

```

RESULT 10
US-09-784-249-2
; Sequence 2, Application US/09784249
; Patent No. US20010027184A1
; GENERAL INFORMATION:
; APPLICANT: Sanjay Kumar
; APPLICANT: Cheng Zou
; TITLE OF INVENTION: SPRINE/THPEONINE PROTEIN KINASE (H-SGK2)
; FILE REFERENCE: GH-70124-C1
; CURRENT APPLICATION NUMBER: US/09/784,249
; CURRENT FILING DATE: 2001-02-15
; PRIOR APPLICATION NUMBER: 09/997,212
; PRIOR FILING DATE: 1997-12-23
; PRIOR APPLICATION NUMBER: 60/051,446
; PRIOR FILING DATE: 1997-07-01
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 496
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
US-09-784-249-2

```

```

Query Match          5.7%; Score 93.5; DB 10; Length 496;
Best Local Similarity 31.0%; Pred. No. 1.1;
Matches 27; Conservative 17; Mismatches 36; Indels 7; Gaps 3;

```

```

QY 80 IASAPTEFPVSVVYVQIIVITGSGFDNNKAVLEPPYSDFPKLCKALLKTFREIEDEVE 139
DB 18 IIPSDHEHREKKRPVYKIV-----SVGRSEMFVPRVAEFDKLYNTLKQF PAMALK 71
QY 140 FPRKHITG-NFAEEMICEPPALQEVLT 165
DB 72 IIPKPIFGNFDPPFIKQPPAGLNFTI 98

```

```

RESULT 11
US-09-910-150-2
; Sequence 2, Application US/09910150
; Patent No. US2002006869A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Meyers, Rachel
; APPLICANT: Rudolph-Owen, Laura
; APPLICANT: Kapeller-Liberman, Rosana
; APPLICANT: Tsai, Fong Ying
; TITLE OF INVENTION: 1327, 14489, 2245 OR 1628 NOVEL HUMAN
; TITLE OF INVENTION: 1327, 14489, 2245 OR 1628 NOVEL HUMAN
; FILE REFERENCE: 18155-20020.00

```

```

; CURRENT APPLICATION NUMBER: US/09/910,150
; CURRENT FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: US 60/219,028
; PRIOR FILING DATE: 2000-07-18
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1066
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-910-150-2

```

```

Query Match          5.6%; Score 92.5; DB 10; Length 1066;
Best Local Similarity 26.9%; Pred. No. 4.6;
Matches 28; Conservative 21; Mismatches 48; Indels 7; Gaps 2;

```

```

QY 87 EFPVSEFVVYQ IIVITGSGFDNNKAVLEPPYSDFPKLCKALLKTFREIEDEVE 139
DB 21 QHPRGYTYVKYTVARVVRPNPNDVQETIWKRYSDPKLHKELWQJHKLFRHSELFP 80
QY 140 FPRKHITG-NFAEEMICEPPALQEVLT 165
DB 81 FAKGIVFGPFDETVIEPPQCAEDLDGFSANIPALVNSKQLEDF 124

```

```

RESULT 12
US-09-815-242-11828
; Sequence 11828, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Chisoe, Fari L.
; APPLICANT: Zykind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA-011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11828
; LENGTH: 1162
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-11828

```

```

Query Match          5.4%; Score 89; DB 10; Length 1162;
Best Local Similarity 23.6%; Pred. No. 11;
Matches 60; Conservative 30; Mismatches 80; Indels 84; Gaps 13;

```

```

QY 106 FPNHVALEPPYSDFPKLCKALLKTFREIEDEVE 145
DB 84 FDKAEITLVGEVAGYAEISTFPVSPGPPQGTYYLTPYRPPFTTSTGTCGDRSCT 143

```



QY 146 -----TGNFAEE-----MICEPRPALOEVLGLYALRACVRSREFLDEL 184  
DB 144 IEQGMISKLEIARPELDRFIEBAGISKYKRRPETSRR-----TRTQENLARI 194  
QY 185 TRPELFEAAGCIRAGGYPAIEELLPLVPLQCKITAFCPAAVPAALCAVL--LCHPDIDR 242  
DB 195 T-DLREELG-----FQLEPLHPQASAEKQOEH--KAEEROLKQOLGAVWRDLNE 242  
QY 243 PA-----EAFAGER-----ALQRLQAREGHRVYAPLLDMVLAIALGKDF 284  
DB 243 QVQGRERVIGDEIAEFALVAEORGADAGIERF--PDGHHELEFENQVQARFYSVGGDI 300  
QY 285 VTQERLEESQLR 298  
DB 301 ARVEGSIHQGQORQ 314

RESULT 13  
US-09-815-242-10617  
; Sequence 10617, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haseibeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; PRIORITY FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 10617  
; LENGTH: 1045  
; TYPE: PRT  
; ORGANISM: Enterococcus faecalis  
US-09-815-242-10617

Query Match 5.3%; Score 88; DB 10; Length 1045;  
Best Local Similarity 21.0%; Pred. No. 12;  
Matches 61; Conservative 57; Mismatches 128; Indels 44; Gaps 11;  
QY 48 GLSSNSMTTRELQYQWONQKCFW-KHYVLLFETASARIEPVS-KFVVYQITVIGTGS 105  
DB 412 GTLEFANFECSSVADHWQHFVEFWQFNQAWQKISQNGVELHELTORPAVQO-----QOKS 487  
QY 106 FDNKAVLERRYSDFAKLQKALKTFRFEEI-----DVEPRKHLGNP-----AAE 152  
DB 488 AE-EKKLQTKSSQWMSLDIQRLSLLEBGEPCVCSLEHPKQOTHOEVSLEIDQAE 545  
QY 153 MICEPRPALOEVLGLYALRACVRSREFLDELTPPELFEAAGCIRAGGYPAIEELLPLV 212  
DB 546 ELTEVEKTVQRFTELTLALGAKEKQOE-----SQLOEQ-----EAATEBQBLAAQFA 594

QY 213 PLQCKITAFCPAAVPAALCAVLCHRDLPRAEFAAGERALOR-----LQAREGHR 265  
DB 595 DQPLPLTGLTFSQVTPAIEAVESQLAKEKQOTQKULTEISVSKDRLAELDEQVANSOR 654  
QY 266 YAPLLDMVLAIALGKDFVTQERLEESQLRPPTPGITLKEITREYL 315  
DB 655 FVLAHQVETWQOSLER--ITQOQMIASQLDLAT--VTEEMTQOAL 699

RESULT 14  
US-10-098-979-2  
; Sequence 2, Application US/10098979  
; Patent No. US2002017207A1  
; GENERAL INFORMATION:  
; APPLICANT: Myriad Genetics, Incorporated  
; APPLICANT: Sugiyama, Janice  
; APPLICANT: Cimbora, Daniel  
; TITLE OF INVENTION: TSG101-INTERACTING PROTEINS AND USE THEREOF  
; FILE REFERENCE: 1907.06  
; CURRENT APPLICATION NUMBER: US/10/098,979  
; CURRENT FILING DATE: 2002-03-14  
; PRIOR APPLICATION NUMBER: US 60/276,259  
; PRIOR FILING DATE: 2001-03-14  
; PRIOR APPLICATION NUMBER: US 60/304,101  
; PRIOR FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: to be assigned  
; PRIOR FILING DATE: 2001-10-22  
; PRIOR APPLICATION NUMBER: to be assigned  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 1993  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-098-979-2

Query Match 5.3%; Score 87.5; DB 9; Length 1993;  
Best Local Similarity 22.4%; Pred. No. 33;  
Matches 72; Conservative 42; Mismatches 109; Indels 99; Gaps 17;  
QY 49 LSNSMTTRELQYQWONQK-----CRWKHVKLFE-IASARIEER-KVSK 92  
DB 292 VSSIMSTLRELQGTWNLDMVGQKLILKSVLDQSSHKVAFDKNSTLMARSLSR 351  
QY 93 FVY-----YQITVIGTGSFNNKAVL---ERRYSDFAKLQKALK-----TFREIED 137  
DB 352 FLLTGSLEAVQVQVNDLQNLDDLEKQERSLOKFGSITNQLKECHPVTETLTNTLKE 411  
QY 138 VEPFKHLTGNFAEEMICERRAL-----QEYLGILYAIRGV-----RSPFLDFITP 186  
DB 412 VNMKNMLLEBIAEOL--QSSKALQIMORYKD--YSKQCASTVQOCEPRTNELLAATN 467  
QY 187 P-----EIRFAFGCIRAGGYPAIEELLPLVPLQCKITAFCPAAVPA-- 229  
DB 468 KDIADEVAWTIQDNCNDLKLGLTGVK-----DSLFTLHEIGEOLKQOVASASASAO 519  
QY 230 -----LCAV--LCHRDLPRA-----EAFAGERALQF--LQAREGHRVYAPLLD 271  
DB 520 SDQSLSGHLCALBQALCKQOTSLQAGVLDYETFAKSLALEMIVAEELIQQGCP-- 576  
QY 272 AMVRLAVALGKDFVTQERLE 293  
DB 577 -----SHSSDLSTIGERME 591

RESULT 15  
US-10-042-417-28  
; Sequence 28, Application US/10042417  
; Patent No. US2002013082A1  
; GENERAL INFORMATION:  
; APPLICANT: Pagano, M.  
; TITLE OF INVENTION: METHODS TO IDENTIFY COMPONENTS USEFUL FOR THE TREATMENT OF

; TITLE OF INVENTION: PROLIFERATIVE AND DIFFERENTIATIVE DISORDERS  
 ; FILE REFERENCE: 5914-090-999  
 ; CURRENT APPLICATION NUMBER: US/10/042,417  
 ; CURRENT FILING DATE: 2002-01-07  
 ; PRIOR APPLICATION NUMBER: 60/260,179  
 ; PRIOR FILING DATE: 2001-01-05  
 ; NUMBER OF SEQ ID NOS: 89  
 ; SOFTWARE: Patent Ver. 2.0  
 ; SEQ ID NO: 28  
 ; LENGTH: 621  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US: 10-042 417-28

Query Match 5 2%; Score 85.5; DB 12; Length 621;

Best Local Similarity 25.4%; Pred. No. 10; Mismatches 80; Indels 79; Gaps 18;

Matches 67; Conservative 38; Mismatches 80; Indels 79; Gaps 18;

QY 48 GLSSNSMTRELOQ---YMONQ-KCPW---KHVTLFEIASAPIEEYVSKFVYQI 98  
 DB 53 GRVSTGRLRELQSSGKWKKEQFVRWMSLMKHYSPTDYV--NMLEEYKV----- 102  
 QY 99 IVLTGSHUNNKAVLEPPYSTFAKIQKALKTPEE-----IEDVEPPRHLTGNF 149  
 DB 103 -----RQKGLER---KIVASFSEKFESEHVPDNCGFSDIENLEGPETF----F 144  
 QY 150 AEEMIC---EPPALQ-EYLG--LLYAIPCVPSPREFLDPLTPRELREAF--GCLRAGQ 200  
 DB 145 EDELVCLINMEGRKALTWKYAF::YYLPQKILNNLFAFLQCPDYESYLEGAVYIDQ 204  
 QY 201 YPRAL-ELLRLVLPLOEKLTAHCPAAVPAALCAVLLCHRDLD--RPAAFAAGERAL--- 254  
 DB 205 YCNPLNDISLKDIOAQID-----SIVELVYTL--PGINEPHRSLAFKAGESSMIME 254  
 QY 255 QRIQAREGHRYVAPLIDAMVRLAY 278  
 DB 255 TELQSQ-----VLDAMVYLVLY 270

Search completed: April 20, 2003, 14 12 15  
 Job time : 44 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Comugen Ltd.

# OM protein - protein search, using sw model

Run on: April 20, 2003, 14:04:41 ; Search time 15 Seconds  
(without alignments) 619.843 Million cell updates/sec

Title: US-09-816-697A-2  
Perfect score: 1650  
Sequence: 1 MASPEHPGSGMPTTCT... PPTPTGTLKELTVFYL 316

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database: Issued Patents AA:  
1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/1aa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/1aa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	99.5	6.0	696	US-08-906-865-4	Sequence 4, Appl
2	99.5	6.0	696	US-08-129-669-4	Sequence 4, Appl
3	95.5	5.8	2930	US-09-417-842-2	Sequence 2, Appl
4	90	5.5	3072	US-09-413-814-93	Sequence 3, Appl
5	90	5.5	3079	US-09-413-814-80	Sequence 80, Appl
6	85.5	5.2	522	US-08-625-322-2	Sequence 2, Appl
7	85	5.2	338	US-08-820-719-12	Sequence 12, Appl
8	84	5.1	486	US-08-821-355A-8	Sequence 8, Appl
9	84	5.1	486	US-09-003-687A-8	Sequence 8, Appl
10	84	5.1	486	US-09-136-605-8	Sequence 8, Appl
11	81	4.9	451	US-08-625-322-4	Sequence 4, Appl
12	81	4.9	297	US-08-622-352A-7	Sequence 7, Appl
13	81	4.9	287	US-08-622-352A-9	Sequence 9, Appl
14	81	4.9	287	US-08-826-390-9	Sequence 9, Appl
15	81	4.9	816	US-08-533-306A-6	Sequence 6, Appl
16	81	4.9	816	US-08-742-923A-6	Sequence 6, Appl
17	81	4.9	885	US-08-533-306A-4	Sequence 4, Appl
18	81	4.9	995	US-08-742-923A-4	Sequence 4, Appl
19	80.5	4.8	366	US-09-210-843-2	Sequence 2, Appl
20	80	4.8	723	US-09-434-408-2	Sequence 2, Appl
21	80	4.8	1263	US-09-446-504-6	Sequence 6, Appl
22	80	4.8	1263	US-09-712-266-6	Sequence 6, Appl
23	80	4.8	1263	US-09-091-889A-4	Sequence 4, Appl
24	79	4.8	1580	US-08-804-198-5	Sequence 5, Appl
25	78.5	4.8	706	US-08-804-198-5	Sequence 5, Appl
26	78.5	4.8	706	US-08-339-152A-29	Sequence 29, Appl
27	78.5	4.8	706	US-08-007-999-4	Sequence 4, Appl

28	78.5	4.8	706	US-08-689-276A-4	Sequence 4, Appl
29	78	4.7	993	US-09-060-410-4	Sequence 4, Appl
30	78	4.7	3111	US-08-460-309-4	Sequence 4, Appl
31	78	4.7	3111	US-08-125-077-4	Sequence 4, Appl
32	76.5	4.6	557	US-09-027-064-2	Sequence 2, Appl
33	76.5	4.6	557	US-09-271-815-2	Sequence 2, Appl
34	76.5	4.6	2101	US-08-466-380-4	Sequence 4, Appl
35	76.5	4.6	2101	US-08-470-950-4	Sequence 4, Appl
36	76.5	4.6	2101	US-08-467-781-4	Sequence 4, Appl
37	76.5	4.6	2101	US-08-195-487-4	Sequence 4, Appl
38	76.5	4.6	2101	US-08-483-924-4	Sequence 4, Appl
39	76.5	4.6	2101	US-09-452-294-1	Sequence 1, Appl
40	76.5	4.6	2101	PCT US03-06160-4	Sequence 4, Appl
41	76	4.6	454	US-08-764-870-5	Sequence 5, Appl
42	76	4.6	454	US-08-940-115-6	Sequence 5, Appl
43	76	4.6	454	US-08-940-115-6	Sequence 5, Appl
44	76	4.6	2639	US-09-080-983-3	Sequence 3, Appl
45	75.5	4.6	204	US-08-419-414-9	Sequence 9, Appl

## ALIGNMENTS

RESULT 1  
US-08-906-865-4  
Sequence 4, Application US/08906865  
Patent No. 6040168  
GENERAL INFORMATION:  
APPLICANT: Greengard, Paul  
APPLICANT: Forton, Barbara  
APPLICANT: Kao, Hung-Teh  
TITLE OF INVENTION: DNA ENCODING THE HUMAN SYNAPSIN III GENE  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: David A. Jackson, Esq.  
STREET: 411 Hackensack Ave, Continental Plaza, 4th  
FLOOR  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/906,865  
FILING DATE:  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-202  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-487-5800  
TELEFAX: 201-343-1684  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 696 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
DESCRIPTION: /desc = "Synapsin Ia"  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
US-08-906-865-4  
Query Match 6.0% Score: 99.5; PR 3; Length: 696;  
Best Local Similarity 25.0% Pred. No. 0.031;

Matches 72; Conservative 29; Mismatches 96; Indels 91; Gaps 17;

QY 4 PEPHG--SPGCM-GPITQCTARTQCEAPATGDDLPHPGPDGHLDTHTSGLSNSSMTTEL 60  
 Db 32 PPPGAISSPGATPGATGATERSGGVAPAPASPAASPGSGGGGFFSSLSNAVKQT--- 88  
 QY 61 QQWQWQKCPWKHVKLLFFIASAPTEEPVSKVYVQIIVITQSGFENKA- -V 112  
 Db 89 -----AAAATFSEQ-----VGGSGGAGRGGAASRVLLV 118  
 QY 113 IEPYSDFAKLCALKTPEPEI---PDVFPP -KHLTGNFAEMICEPPALQCY 164  
 Db 119 IDEPHITMAKYEFG--KIHGEIDIKVQAEFSDNLVHNAKGFSVMEVLRN----- 176  
 QY 165 IGLIVAIKCPSPSPFLIDF-LTPPELIPAPAGTLPACQYPALELLLPVLPQKLTANCP 223  
 Db 171 ---GVYVPSLKP--DEVLIP---QHAFSMAFNQDY RSL VIGQ----- 206  
 QY 224 AAAPALCAVILCPDIDPPAEVFAAGCEPALQPIGAPEGHPYVADLLD 271  
 Db 207 YACIPVNSLHSVYNPCDKE-WVFAQMVRLHKLTGEE- -FLIID 248

RESULT 2  
 US-09-129-668 4  
 Sequence 4, Application US/09129668B  
 Patent No. 6429010  
 GENERAL INFORMATION:  
 APPLICANT: Greengard, Paul  
 APPLICANT: Porton, Barbara  
 APPLICANT: Kao, Hung Teh  
 TITLE OF INVENTION: DNA ENCODING THE HUMAN SYNAPSIN III GENE AND USES  
 TITLE OF INVENTION: THEREOF  
 FILE REFERENCE: 600-1-202 CIP  
 CURRENT APPLICATION NUMBER: US/09/129,668B  
 CURRENT FILING DATE: 1998-08-05  
 EARLIER APPLICATION NUMBER: 08/906,865  
 EARLIER FILING DATE: 1997-09-06  
 NUMBER OF SEQ ID NOS: 8  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO 4  
 LENGTH: 696  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-09-129-668 4

Query Match 6.0%; Score 99.5; DB 4; Length 696;  
 Best Local Similarity 25.0%; Pred. No. 0.031; Indels 91; Gaps 17;  
 Matches 72; Conservative 29; Mismatches 96; Indels 91; Gaps 17;

QY 4 PEPHG--SPGCM-GPITQCTARTQCEAPATGDDLPHPGPDGHLDTHTSGLSNSSMTTEL 60  
 Db 32 PPPGAISSPGATPGATGATERSGGVAPAPASPAASPGSGGGGFFSSLSNAVKQT--- 88  
 QY 61 QQWQWQKCPWKHVKLLFFIASAPTEEPVSKVYVQIIVITQSGFENKA- -V 112  
 Db 89 -----AAAATFSEQ-----VGGSGGAGRGGAASRVLLV 118  
 QY 113 IEPYSDFAKLCALKTPEPEI---EDVEFPP--KHLTGNFAEMICEPPALQCY 164  
 Db 119 IDEPHITMAKYEFG--KIHGEIDIKVQAEFSDNLVHNAKGFSVMEVLRN----- 176  
 QY 165 IGLIVAIKCPSPSPFLIDF-LTPPELIPAPAGTLPACQYPALELLLPVLPQKLTANCP 223  
 Db 171 ---GVYVPSLKP--DEVLIP---QHAFSMAFNQDY RSL-----VIGQ----- 206  
 QY 224 AAAPALCAVILCPDIDPPAEVFAAGCEPALQPIGAPEGHPYVADLLD 271  
 Db 207 YACIPVNSLHSVYNPCDKE-WVFAQMVRLHKLTGEE- -FLIID 248

RESULT 3  
 US-09-417-822 2

Sequence 2, Application US/09417822  
 Patent No. 6344549  
 GENERAL INFORMATION:  
 APPLICANT: Keegan, Kathy  
 TITLE OF INVENTION: ATR-2  
 FILE REFERENCE: 27866/35633  
 CURRENT APPLICATION NUMBER: US/09/417,822  
 CURRENT FILING DATE: 1999-10-14  
 NUMBER OF SEQ ID NOS: 43  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO 2  
 LENGTH: 2930  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-09-417-822 2

Query Match 5.8%; Score 95.5; DB 4; Length 2930;  
 Best Local Similarity 22.7%; Pred. No. 0.81;  
 Matches 70; Conservative 48; Mismatches 112; Indels 79; Gaps 15;

QY 48 GLSNSSMTTELQYVWQWQKCPWKHVKLLFFIASAPTEEPVSKVYVQIIVITQ 103  
 Db 1711 GQAAESKQSKPEMP-----ETPSPSPVARTYVWFPNDEMVLVLPKLD 1758  
 QY 104 GSEFNNKAVLEPYSDFAKLCALUK--TPPRKIFVAPPPKULTANFAPE-MI'EPRA 160  
 Db 1759 GSDDEVLS-LQEGLTIVKLGKLEEFLEFGABSVVHSHLQHYSHHTQULQVQRA 1817  
 QY 161 LQEVLLVYAIKCPSPSPFLIDF-LTPPELIPAPAGTLPACQYPALELLLPVLPQKLTANCP 220  
 Db 1818 VQEAIVQ-----KLNFEQWIT--HYCAFNNLBATQ-----LASTQISTQMDLG- 1862  
 QY 221 HPPAAVPAATF-----AVLQHPRLCPAPAPAGCEPALQPIGAPEGHPYVADLLD 271  
 Db 1863 --PSPVPTATFLQNAQAHLSQ'EQ--LQGEVVALUQPPSVLKQ'LEQLHHV----- 1913  
 QY 272 AMVRLAYALG-----KCPVTLQEFLL-----EESQLPFFTP EGIT 305  
 Db 1914 ATVALQYPALEKQKHPTEGWTWMPRLICNTTVERQYELVPKYEMVYAP'PPPTVQCFIT 1973  
 QY 306 LKELTYREY 314  
 Db 1974 ATEMQLQRY 1982

RESULT 4  
 US-09-413-814-93  
 Sequence 93, Application US/09413814  
 Patent No. 6225064  
 GENERAL INFORMATION:  
 APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH  
 APPLICANT: Bristol-Myers Squibb, Co.  
 APPLICANT: Bayer, Stefan  
 APPLICANT: Bloeker, Helmut  
 APPLICANT: Brandt, Petra  
 APPLICANT: Cino, Paul M  
 APPLICANT: Dougherty, Brian A  
 APPLICANT: Goldberg, Steven L  
 APPLICANT: Hoeller, Gerhard  
 APPLICANT: Hoeller, Joachim  
 APPLICANT: Reichenbach, Hans  
 TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polypeptide or  
 TITLE OF INVENTION: heteropolypeptide compounds  
 FILE REFERENCE: PCT/US 99/23535  
 CURRENT APPLICATION NUMBER: US/09/413,814  
 CURRENT FILING DATE: 1999-10-07  
 EARLIER APPLICATION NUMBER: DE 198 46 493.2  
 EARLIER FILING DATE: 1998-10-09  
 NUMBER OF SEQ ID NOS: 107  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 93  
 LENGTH: 3072  
 TYPE: PRT





APPLICANT: Clevers, Hans  
APPLICANT: Korinek, Vladimir  
APPLICANT: Morin, Patricia  
APPLICANT: Kinzler, Kenneth  
APPLICANT: Vogelstein, Bert  
APPLICANT: Spake, Andrew  
TITLE OF INVENTION: Beta Catenin, TCF-4, and APC  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Banner & Witcoff, Ltd.  
STREET: 1001 G Street, N.W.  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20001  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/003,687A  
CLASSIFICATION:  
FILING DATE:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: 08/821,355  
FILING DATE: 20-MAR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Kagan, Sarah A  
REGISTRATION NUMBER: 32,145  
REFERENCE/DOCKET NUMBER: 1107.05064  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
TELEX: 97430 BMB UT  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 486 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: No. 5998600e  
US-09-003-687A-8

Query Match  
Best Local Similarity 5.1%, Score 84, DB 2, Length 486;  
Matches 69, Conservative 35, Mismatches 95, Indels 114, Gaps 16,

QY 4 FEHPSPGCMGPIITQCTARTQOEAPATGPDLPHP-----GPDGH-LDTHSGLS 50  
DB 215 FLYFLSP-----SCGYOHFAFPAAGAYPRFTHSLSMLGSGVGHRAIPRAIV 267  
QY 51 SNSSMTPRELQYWNQCPKPMKHVLLFEIASAPIE-----ERKYSKFVYV---QI 98  
DB 268 PRSG--KQELQPFDRMLTKQ-----AESKAKEKAKPTIKKPLNAFVLKEMBA 315  
QY 99 IVYQTSFQNNKA---VLPPYSDPAKIQKALLKTPPEIEVEFRPPHLTGNFAEMIC 155  
DB 316 KYIAECTLKESAAINQILGPWMAISP-----EEQAYVELARF-----354  
QY 156 ERRRALQEVLYGLVLAIR-----CVPRSPFLDPLTPPELREAFGCLPAGQYPALELLR 210  
DB 355 EROLHMQLYPG--WSAPNNYKPKPPSPRKHQSTTGAKPVAF-----GYP-----399  
QY 211 VLPLEKLTAMCPAAVPAALCA-----VLGHPDLUPPAEFAAGEPALQPLQA 259  
DB 400 -----EKAAPAPFLPMTVLAAPGQPLRPTHHTICCP-----ASPONCLLARS 444  
QY 260 REGHRYAPFLDA 272  
DB 445 RHLHPQVSPLLSA 457

RESULT 10  
US-09-136-605-8  
Sequence 8, Application US/09136605A  
Patent No. 6140052  
GENERAL INFORMATION:  
APPLICANT: He, Tong-Chuan  
APPLICANT: Kinzler, Kenneth  
APPLICANT: Vogelstein, Bert  
TITLE OF INVENTION: Beta Catenin, TCF-4, and APC Interact to  
FILE REFERENCE: 1107.75741  
CURRENT APPLICATION NUMBER: US/09/136,605A  
CURRENT FILING DATE: 1998-08-20  
EARLIER APPLICATION NUMBER: 08/821,355  
EARLIER FILING DATE: 1997-03-20  
EARLIER APPLICATION NUMBER: 09/003,687  
NUMBER OF SEQ ID NOS: 28  
SOFTWARE: FASTSEQ for Windows Version 3.0  
SEQ ID NO 8  
LENGTH: 486  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-136-605-8

Query Match  
Best Local Similarity 5.1%, Score 84, DB 4, Length 486;  
Matches 69, Conservative 35, Mismatches 95, Indels 114, Gaps 16;

QY 4 FEHPSPGCMGPIITQCTARTQOEAPATGPDLPHP-----GPDGH-LDTHSGLS 50  
DB 215 FLYFLSP-----SCGYOHFAFPAAGAYPRFTHSLSMLGSGVGHRAIPRAIV 267  
QY 51 SNSSMTPRELQYWNQCPKPMKHVLLFEIASAPIE-----ERKYSKFVYV---QI 98  
DB 268 PRSG--KQELQPFDRMLTKQ-----AESKAFFFAVYPTIKKPLNAFVLKEMBA 315  
QY 99 IVYQTSFQNNKA---VLPPYSDPAKIQKALLKTPPEIEVEFRPPHLTGNFAEMIC 155  
DB 316 KYIAECTLKESAAINQILGPWMAISP-----EEQAYVELARF-----354  
QY 156 ERRRALQEVLYGLVLAIR-----CVPRSPFLDPLTPPELREAFGCLPAGQYPALELLR 210  
DB 355 EROLHMQLYPG--WSAPNNYKPKPPSPRKHQSTTGAKPVAF-----GYP-----399  
QY 211 VLPLEKLTAMCPAAVPAALCA-----VLGHPDLUPPAEFAAGEPALQPLQA 259  
DB 400 -----EKAAPAPFLPMTVLAAPGQPLRPTHHTICCP-----ASPONCLLARS 444  
QY 260 REGHRYAPFLDA 272  
DB 445 RHLHPQVSPLLSA 457

RESULT 11  
US-09-625-322-4  
Sequence 4, Application US/08625322  
Patent No. 5804412  
GENERAL INFORMATION:  
APPLICANT: Gill, Gordon N.  
APPLICANT: Kurihara, Richard C.  
APPLICANT: Cadena, Deborah L.  
TITLE OF INVENTION: Sorting Nexins and Methods of Using Same  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Campbell and Flores  
STREET: 4370 La Jolla Village Drive, Suite 700  
CITY: San Diego  
STATE: California  
COUNTRY: USA  
ZIP: 92122  
COMPUTER READABLE FORM:

```

MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/625,322
FILING DATE: 01-APR-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-UD 1955
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 451 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-625-322-4

```

```

Query Match 4.9%; Score 81.5; DB 1; Length 451;
Best Local Similarity 19.9%; Pred. No. 1.7;
Matches 60; Conservative 45; Mismatches 96; Indels 101; Gaps 12.

```

```

QY 77 LFTIASKRIIEERV-----SKFVYVYCIIVITQTSFNNKAVLEPPSPAPKICVALLKTFPP 132
DB 72 IPIEIGVSVHEFVGMQMAVMYRVTTKTSLSMFSKSEFVSNEFTDLGHTTLPITYL 131
QY 133 EELIEDV EFRKRLTG ... HFAEMICEPPRAICGYLGLVAIFCVPPSPFEFL 181
DB 132 HVIIVATSRKSIQMTIVYVGRKSSSTFEVERPRALERYL... QNTVHP 182
QY 182 DEITPELPAPGCLPACQYPPA ... LELLFVLFLCEKLT ... 239
DB 183 TILQDPLRQ---FLRSSELPRAVNTQALSGAGILFMVKKADAVNKKTIKNESDAWFE 239
QY 220 ... AHCVAAPVPAICVALLGHR DLDIFAFA... -----AG 250
DB 240 EFGGEEENILGELKRLVSVKA ... LVCHRELGANTAFKSAAMCENSEDTAL 292
QY 251 ERALQRL-----QAFEGHPYVAPLDMVRLAVALGDF--VTLGERLEES 294
DB 293 SPANISGLAEVEERIKDQHGFQAFAPFYMESPILTSQYIPLIAVGVVFCRHPYCWYWDNA 352
QY 295 QL 296
DB 353 QI 354

```

```

RESULT 12
US-08-622-353-7
; Sequence 7, Application US/08622353
; Patent No. 5700925
; GENERAL INFORMATION:
; APPLICANT: Bishai, William R.
; APPLICANT: Young, Douglas B.
; APPLICANT: Zhang, Ying
; APPLICANT: Demajo, James
; TITLE OF INVENTION: A STATIONARY PHASE, STRESS RESPONSE
; TITLE OF INVENTION: SIGMA FACTOR FROM MYCOBACTERIUM TUBERCULOSIS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Allegretti, LTD
; STREET: 1001 G Street, eleventh floor
; CITY: NM
; STATE: Washington DC
; COUNTRY: USA
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk

```

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/622,353
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Hoschelt, Dale
REGISTRATION NUMBER: 19090
REFERENCE/DOCKET NUMBER: 3181.51220
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9239
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 287 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ORIGINAL SOURCE:
ORGANISM: Streptomyces coelicolor
US-08-622-353-7

```

```

Query Match 4.3%; Score 81, DB 1, Length 287;
Best Local Similarity 21.2%; Pred. No. 0.97;
Matches 66; Conservative 38; Mismatches 108; Indels 100; Gaps 15;

```

```

QY 3 SPENHASKKMSPIQVQARTQAEAFATGPDLPFGSGDGHDTSS-----GSSN 52
DB 6 AQCAPPAP-----PAQADQAPACQAPAPAPQSKSKATPAALIGVFGELKGIARG 56
QY 53 SSXTTFELQVQWQNGPCWHVFF... IETACAFIEEPVYFVVVY... IVL 103 203
DB 57 TP---EHQPVVALIEALFLVAYMAAFPCFHPMIVVWSTIILINALIK 106
QY 106 FNNKAVLEPPYSPFA--KLOKALKTPREIEDEVEPPRK-----HLGNFA 150
DB 107 FDPKRV---QPTFAMTTVGSEIKRYFDQVTVVHVPKRLHEIMVUNSATDITTAFG 163
QY 151 EE---MICEPPRAICGYLGLVAIFVPPSPFEIETFEPELEAFGLPAQYPPALH 206
DB 164 RAPTAEIAEPLIFEFV---LSCTEACRSY-----HATSELAQGEFGCLP 207
QY 207 ILIPVLPQCELTANCPAALVPAICVALLCHRPCLDRPAEFAAGFPALQPCAFEGH... 253
DB 208 GILDRIGYED-----PALNGVE--HPDLV-----HILVQLPEREQPIIL 245
QY 264 -RYVAPLDMV 274
DB 246 LRYSNLTQSOI 257

```

```

RESULT 13
US-08-622-352A-9
; Sequence 9, Application US/08622352A
; Patent No. 5824546
; GENERAL INFORMATION:
; APPLICANT: Bishai, William R.
; APPLICANT: Demajo, James
; TITLE OF INVENTION: REGULATION OF A SIGMA FACTOR
; TITLE OF INVENTION: FROM MYCOBACTERIUM TUBERCULOSIS
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pillsbury Madison & Sutro, L.L.P.
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk

```



```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/622,352A
; FILING DATE: 27-MAR-1996
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 287 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptomyces coelicolor
;
US-08-622-352A-9

```

```

Query Match          4.9% Score 81; DB 2, Length 287,
Best Local Similarity 21.2%; Pred. No. 0.97;
Matches 66; Conservative 38; Mismatches 108; Indels 100; Gaps 15;

```

```

QY 3 SPHPSPCGMPITQCTAPTOGEAPATGRLPHGPDGHLDTHS-----GLSSN 52
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   6 AQAAPPAP-----PAQAQAQAPQAQAQAPAPQPSGADTRALTQVLFGEKSLAPG 56
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   53 SMTRELCQYWGNCQCPKHWKL---LFEIASARIEERKVSKEFYVQI---IVICTGS 105
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   57 TP-----EHDVRAALIEANLPIVRYAARFSPSPNEPMEDVQVGTIGLINADR 106
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 106 FDNKVALEPPYSQFA--FLQVALLTFFEEIEDEVEFPK-----HLTGNSFA 150
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 107 FDEPSPV---QPTFAMPTVWGSEIYFPDNPVTIVHPFLHLMVQVNSATEDLTAFG 163
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 151 EE---MICEPPALAPVYGLVAIPQVPPSEFLDTTPPELPAFCCLPAQGVPALE 206
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 164 RSTTAEIAERLITTEEV---LSCIEAGRSY-----HATSLMAQEGGGLP 207
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 207 LILRVFLQCEKLTACPAAVPALCAVLCGRDLFPFAFAAGEPALCPLEAPEGH--- 263
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 208 GLDRLGYED-----PALDGV--HRLVY-----HLVQLPREPQILL 245
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 264 RYAPLIDAMV 274
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 246 LRYSNLTQSQI 257
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

## RESULT 14

```

US-08-826-390-9
; Sequence 9, Application US/08826390
; Patent No. 6004764
; GENERAL INFORMATION:
; APPLICANT: Bishai, William R.
; APPLICANT: Young, Douglas B.
; APPLICANT: Zhang, Yang
; APPLICANT: Demalo, James
; TITLE OF INVENTION: Stationary Phase Stress Response
; TITLE OF INVENTION: Sigma Factor From Mycobacterium tuberculosis, and
; TITLE OF INVENTION: Regulation Thereof
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pillsbury Madison & Sutro, L.L.P.
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect
; CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/08/826,390
; FILING DATE: 27-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/622,353
; FILING DATE: 27-MAR-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/622,352
; FILING DATE: 27-MAR-1996
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 287 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptomyces coelicolor
;
US-08-826-390-9

```

```

Query Match          4.9% Score 81; DB 3; Length 287,
Best Local Similarity 21.2%; Pred. No. 0.97;
Matches 66; Conservative 38; Mismatches 108; Indels 100; Gaps 15;

```

```

QY 3 SPHPSPCGMPITQCTAPTOGEAPATGRLPHGPDGHLDTHS-----GLSSN 52
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   6 AQAAPPAP-----PAQAQAQAPQAQAQAPAPQPSGADTRALTQVLFGEKSLAPG 56
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   53 SMTRELCQYWGNCQCPKHWKL---LFEIASARIEERKVSKEFYVQI---IVICTGS 105
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   57 TP-----EHDVRAALIEANLPIVRYAARFSPSPNEPMEDVQVGTIGLINADR 106
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 106 FDNKVALEPPYSQFA--FLQVALLTFFEEIEDEVEFPK-----HLTGNSFA 150
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 107 FDEPSPV---QPTFAMPTVWGSEIYFPDNPVTIVHPFLHLMVQVNSATEDLTAFG 163
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 151 EE---MICEPPALAPVYGLVAIPQVPPSEFLDTTPPELPAFCCLPAQGVPALE 206
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 164 RSTTAEIAERLITTEEV---LSCIEAGRSY-----HATSLMAQEGGGLP 207
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 207 LILRVFLQCEKLTACPAAVPALCAVLCGRDLFPFAFAAGEPALCPLEAPEGH--- 263
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 208 GLDRLGYED-----PALDGV--HRLVY-----HLVQLPREPQILL 245
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 264 RYAPLIDAMV 274
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 246 LRYSNLTQSQI 257
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

## RESULT 15

```

US-08-533-306A-6
; Sequence 6, Application US/08533306A
; Patent No. 5837457
; GENERAL INFORMATION:
; APPLICANT: Liu, Pu
; APPLICANT: Collins, Francis S.
; APPLICANT: Siciliano, Michael J.
; APPLICANT: Claxton, David
; TITLE OF INVENTION: Markers for Detection of Chromosome 16
; TITLE OF INVENTION: Rearrangements
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
; STREET: P.O. Box 828
; CITY: Bloomfield Hills
; STATE: MI
; COUNTRY: USA
; ZIP: 48303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

```

## CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/533,306A  
 FILING DATE: September 25, 1995  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Smith, DeAnn F.  
 REGISTRATION NUMBER: 36683  
 REFERENCE/DOCKET NUMBER: 2115-00869COB  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (810) 641-1600  
 TELEFAX: (810) 641-0270  
 INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 816 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US: 08-533 306A-6

## Query Match

4.9%; Score 81; DB 2; Length 816;

Post local Similarity 21.9%; Pred. No. 4.9; Mismatches 133; Indels 132; Gaps 13;

Matches 80; Conservative 31;

QY	22	RTQGEAPATGPDLPHPGPDGHLDTSGLSNSMTTRELQGYWQK-----	68
DB	302	RLQGEIIDLVD-----LDNQPQIVSNLEKQPKPFDQLAEFNISRYANPDRA	352
QY	69	CPKKNVYILF-----FIASRIEERKVSFVVYQIVIQTGSFDNNKAVLERISD	119
DB	353	EAAPEKFTFALSLAPALFPALEAELEETNYMLAEEMEDLVSSKDDVGNVHELEKSK	412
QY	120	FA-KLQKALKTFREIED-----VEPRKHILTGNAFEMIC-----ERRAL	161
DB	413	RALEFQMEEMKTLQLEFLLFDFIQAISPAKLPLVNMQNLKQGFEPDLQAPDQNEFRPQL	472
QY	162	QETL-----GLVYIRCVRR-----	176
DB	473	ORQIHVEYTELEDERNERALAAAKKKLEGLKLELQMSALKREFEALIKQLKYLQNM	532
QY	177	SREPLDFTLPBELREAVGCLKAGYPPALFLLFLVPLQEKLTAKGPAAAVPALCNV	233
DB	543	KDQKRLID-APASPRDLVATAKENE-KKAKSLRADLMQLOEDL-----AAABERARKQA	584
QY	234	ILGHEFDIDREPAFAAGCEPALQ-----RLQAREGHEYYAFLLDAMVLAVALGKLEVTIQE	289
DB	585	DLEKFEIAFELASLSLSPNNLICDEPPLEAR-----IAGLEE	621
QY	290	RLFEESQ 295	
DB	622	ELREEQ 627	

Search completed: April 20, 2003, 14:08:21  
 Job time: 21 secs